Performance Guarantees for Physical Layer Authentication in Mission-Critical Communications

HENRIK FORSSELL

Doctoral Thesis in Electrical Engineering
Stockholm, Sweden 2021

KTH Royal Institute of Technology
School of Electrical Engineering and Computer Science
Division of Information Science and Engineering
Malvinas väg 10, 100 44 Stockholm
SWEDEN

TRITA-EECS-AVL-2021:1
ISBN 978-91-7873-727-7

Akademisk avhandling som med tillstånd av Kungl Tekniska högskolan framlägges till offentlig granskning för avläggande av doktorsexamen i Elektroteknik fredagen den 22 Januari 2021 klockan 13:00 i F3, Lindstedtsvägen 26, Stockholm.

© December 2020 Henrik Forssell, unless otherwise noted.

Tryck: Universitetsservice US AB

Abstract

As the application areas for wireless communications are expanding, we also see new security vulnerabilities arise due to the open nature of the wireless medium. One particularly challenging problem is how to guarantee the security of emerging mission-critical communications, e.g., realized by fifth generation (5G) mobile networks, that will enable use-cases like industrial automation, vehicular communications, and smart grids. As the room for security overhead is limited in mission-critical communications, mainly due to the associated strict requirements on latency and reliability, new lightweight security techniques are researched within the area of physical layer security. In particular, feature-based physical layer authentication (PLA), exploiting transmitter-specific features extracted from received signals for device authentication, is considered a promising solution for lightweight authentication and intrusion detection in mission-critical communications. In this thesis, we provide mathematical tools for analyzing channel-based PLA schemes, and in particular, for deriving worst-case performance guarantees appropriate for mission-critical contexts. We consider worst-case performance guarantees for feature-based PLA from two perspectives:

Firstly, we provide mathematical bounds on the delay-performance impacts that arise due to the unlikely but inevitable erroneous authentication decisions (i.e., false alarms and missed detections). We model the PLA scheme using queueing analysis, develop models for active impersonation attacks, and derive bounds on the queueing delay violation probability using tools from stochastic network calculus. We consider the performance for both single- and multiple-antenna receiver architectures, and furthermore, a distributed multiple-antenna system in which we analyze varying degrees of distributed processing. These results establish under which practical deployments and channel conditions feature-based PLA would constitute a viable option for mission-critical applications. For instance, we find that for low-mobility scenarios with line-of-sight conditions, as exemplified by an industrial automation scenario with fixed sensor deployment, PLA can be used for strongly enhanced security while simultaneously maintaining mission-critical latency deadlines with high reliability. Moreover, we discuss extensions that would allow analysis of scenarios without line-of-sight and with higher mobility.

From the second perspective, we provide tools for deriving the worst-case detection performance under optimal attackers that are aware of the PLA scheme. First, we consider a distributed PLA setting where authentication is based on the channel-states observed at multiple distributed radio-heads. We derive the optimal single-antenna attack strategy and corresponding missed detection probability, and provide a heuristic method for finding the optimal spatial attack position with respect to a given deployment. We then extend the results by considering a multiple-antenna attacker, the corresponding optimal pre-coding strategies, and the detection performance under the worst-case attacker. Furthermore, we analyze the impacts of limited channel state information (CSI) and power budgets at the attacker and provide a counter-strategy that can be used by the PLA receiver. With the single-antenna attacker, our results show significant detection performance benefits

from a distributed antenna setting, which argues for practical relevance of PLA within modern 5G technologies like coordinated multi-point (CoMP) and distributed multiple-input multiple-output (MIMO) systems. For the multiple-antenna attacker, we observe significant impacts given perfect CSI knowledge and favorable channel conditions at the attacker. However, under realistic assumptions on power budget, CSI imperfections, and through the proposed counter-strategy, we find that strict detection performance guarantees can be maintained.

Keywords: Physical layer authentication, mission-critical communications, worst-case performance, queueing delay performance, stochastic network calculus, optimal attack strategies.

Sammanfattning

Tillämpningsområdena för trådlös kommunikation expanderar konstant och möjliggör nya applikationer av informationsteknik. Denna utveckling skapar dock samtidigt nya säkerhetsbrister eftersom det trådlösa mediet är öppet för både avlyssning och extern manipulation. Ett viktigt och utmanande problem är hur man kan leverera säkerhetsgarantier för kritisk trådlös kommunikation, som till exempel kan användas för industriell automation, fordonskommunikation, smarta elnät, samt andra applikationer inom femte generationens (5G) mobilnät. Eftersom kritisk trådlös kommunikation karakteriseras av extremt höga krav på latens och pålitlighet har dessa system mycket begränsade resurser för tidskrävande kommunikation och beräkningar. Den senaste forskningen riktar därför bland annat in sig på säkerhetsmetoder i det fysiska kommunikationslagret (PHY-Layer) för att uppnå säker kommunikation utan att överskrida nämnda begränsningar. Autentisering i det fysiska kommunikationslagret är en sådan metod, vilken utnyttjar sändar-specifika egenskaper som kan avläsas från mottagna trådlösa signaler för att verifiera sändarens identitet och detektera potentiella intrång. Denna avhandling utvecklar matematiska verktyg för att analysera kanalbaserad autentisering i det fysiska lagret, med fokus på att härleda prestandagarantier som är lämpliga för kritisk kommunikation. Vi utvecklar sådana garantier utifrån två perspektiv:

För det första tillhandahåller vi matematiskt härleda begränsningar av de fördröjningar som uppstår på grund av de sällsynta men oundvikliga felbeslut som dessa autentiseringsprotokoll resulterar i. Vi modellerar autentiseringsprotokollen med hjälp av köanalys, utvecklar modeller för aktiva impersonationsbaserade attacker samt härleder övre begränsningar för sannolikheten att systemets krävda latens överskrids. Dessa resultat härleds med hjälp av ramverket stochastic network calculus. Analysen utökas från en-antenns mottagare till fler-antennsystem samt ett distribuerat fler-antennsystem med olika grader av distribuerad beslutsfattning. Våra resultat etablerar de praktiska förutsättningar som krävs för att ett kanalbaserat autentiseringprotokoll ska uppfylla de krav som ställs inom kritisk trådlös kommunikation. Resultaten visar att kanalbaserad autentisering, givet ett scenario med låg mobilitet samt direkt siktlinje mellan sändare och mottagare, kan kan användas för förbättrad säkerhet samtidigt som strikta begränsningar på latens upprätthålls. Vidare diskuterar vi möjliga fall under vilka resultaten kan utökas till scenarier med hög mobilitet samt utan direkt siktlinje.

Den andra typen av garantier handlar om att härleda övre begränsningar för detektionsprestandan, i termer av sannolikheten för ett intrång, under optimalt designade attacker. Först studerar vi ett distribuerat autentiseringsprotokoll baserat på kanalobservationer vid flera distribuerade fler-antennsmottagare. Vi härleder den optimala transmissionsstrategin för en angripare med en antenn samt motsvarande sannolikhet för lyckat intrång. För detta fall tillhandahåller vi även en heuristisk metod för att hitta den optimala attackpositionen. Vidare utvecklar vi resultaten till en angripare utrustad med flera antenner, härleder motsvarande optimala strategier samt

detektionsprestandan givet en kompetent angripare med perfekt kanalinformation. Vi analyserar även påverkan av begränsad kanalinformation och effektbegränsningar hos angriparen samt visar en effektiv motstrategi som kan användas av den autentiserande mottagaren. Resultaten visar att en angripare med flera antenner och perfekt kanalinformation kan ha en signifikant påverkan på autentiseringsprestandan. Givet realistiska antaganden om angriparens kanalinformation och effektbudget visar vi dock att säker detektionsprestanda kan garanteras. Resultaten visar även att stora förbättringar erhålls med den distribuerade autentiseringsmetoden, vilket visar praktisk relevans för autentisering i det fysiska lagret inom moderna 5G teknologier så som coordinated multi-point (CoMP) och distribuerade fler-antennsystem.

Acknowledgements

From my perspective, it almost goes without saying that completing this thesis would not have been possible without the help from and continuous dialogue with my supervisors and colleagues. Therefore, there are many people I want to thank:

First and foremost, I want to thank Ragnar Thobaben for giving me the opportunity to pursue my PhD under his supervision. I am always grateful for your support and constant stream of new ideas and perspectives, and I believe this has taught me a lot over these last years. I also want to thank you for always taking time out of a busy schedule to discuss new problems and for encouraging me to continue when things were tough. I want to thank James Gross for all the help and support over these years. Receiving your input on things has always helped me move forward and put my research in a larger context. I also want to thank my former colleague Hussein Al-Zubaidy, who contributed with helpful feedback and discussions during the first part of my PhD.

I want to thank Henrik Sandberg and all the other people in the CERCES project, as well as MSB for funding this project and my PhD position. It was always inspiring to find my work being part of this larger project, and I believe this helped me find new perspectives on my own research and this thesis. I also want to particularly thank my PhD colleagues in the project: Jezdimir, Andreas, Ezzeldin, and Peyiue. I'm happy that I got to know you during this time, and thank you for all the collaborations, lunches, and fikas.

Next, I want to thank Mikael Skoglund and all the past and present colleagues at the ISE division. To all of you: I'm always grateful for having such nice colleagues. I have a few particular people that I feel I need to mention: Thanks to Tobias Oechtering for supervising my master thesis, recommending me for this PhD position, and for always being a nice person to meet in the corridors. I want to thank Mats Bengtsson for doing the formal review of this thesis. I also want to thank my colleague and friend Boules Atef Mouris, whom I shared office with during most of my PhD, and who also kindly helped me with proof-reading this thesis. Thanks to Magnus Jansson for always stepping by my office to check in or talk about marathon training or skiing. Finally, I want to thank Germán Bassi for our teaching collaborations and for the helpful guidance and discussions during the final part of my PhD.

With that said, life is far more than thesis work and research, and I'm very grateful for having such amazing friends outside of KTH as well. You are always there for me when I just need to hang out and have some fun, and I take this opportunity to thank you for this. You know who you are!

Last, but certainly not least, I want to thank my family: My father Clas, my mother Christina, my sister Adina, Anne och Gunnar, and Charlotte with family. You always make me feel supported and loved.

Henrik Forssell,
Stockholm, 24 November 2020

Table of Contents

List of Figures

List of Tables

List of Acronyms

AoA	Angle-of-Arrival
CDMA	Code Division Multiple Access
CN	Connection (MAC Request)
CoMP	Coordinated Multi-Point
CSCG	Circularly Symmetric Complex Gaussian
CSI	Channel State Information
DCN	Disconnection (MAC Request)
DTA	Data (MAC Request)
DTP	Data Transmission Period
FAP	False Alarm Probability
FDMA	Frequency Division Multiple Access
GLRT	Generalized Likelihood Ratio Test
LoS	Line-of-Sight
LTE	Long Term Evolution
MAC	Medium Access Control
MDP	Missed Detection Probability
MIMO	Multiple-Input Multiple-Output
MTC	Machine-Type Communication
OFDM	Orthogonal Frequency Division Multiplexing
PHY	Physical Layer
PLA	Physical Layer Authentication
ROC	Receiver Operating Characteristics
RRH	Remote Radio-Head
RSSI	Received Signal Strength Indicator
SNR	Signal-to-Noise Ratio
SIMO	Single-Input Multiple-Output
TDMA	Time Division Multiple Access
URLLC	Ultra-Reliable Low-Latency Communication

Part I

Thesis Overview

Chapter 1

Introduction

The application areas for digital communication and information processing have rapidly expanded over the previous decades, and we continually see new use-cases emerge. Furthermore, wireless communication technologies extend the possibilities by removing the need for wired connections and enabling remote sensing and control as well as large-scale connectivity between different types of devices and machines. Many of the new applications we see emerging extend beyond the traditional human-centered communication. A decade ago, most wireless devices that were found in wireless local area networks or cellular mobile networks were used almost exclusively for human-centered communication. While enhancement of the user experience in human-centered communications is still demanded, today we see new challenges in developing wireless standards for applications like industrial automation and manufacturing (i.e., Industry 4.0 or industrial internet-of-things), home appliances, vehicular communication, automated healthcare, and smart energy-management systems. One challenge facing these applications is that they often require more timely and reliable delivery of information compared to mobile broadband. This is one reason for the recent research and development towards ultra-reliable and low-latency communications (URLLC), tailored for systems where the often conflicting requirements on reliability and latency are very strict (e.g., packet error rates in the order of $10^{-9} - 10^{-6}$ and latencies below 1 ms). We can expect machine-type wireless applications to become ubiquitous parts of our future society, realizing automated traffic, industries, and general infrastructure. However, from a security perspective, new problems will arise when safety- and security-critical applications become exposed through the wireless medium. Development of both reliable and secure low-latency wireless technologies is therefore one of the most important problems related to future machine-type communications.

The wireless medium is, by its very nature, open to access anywhere in close proximity to the system, and this gives anyone with the right knowledge, hardware, and malicious intentions an opportunity to attack the system. In automated critical infrastructures, the severity and impact of successful attacks can obviously become

catastrophic. Simultaneously, there are many known threats and vulnerabilities associated with modern wireless systems. Broadly speaking, there are passive attacks, like eavesdropping to collect sensitive information, and active attacks, like various forms of denial-of-service attacks, impersonation attacks, and data injection attacks [1]. An attacker can cause denial-of-service in multiple different ways, ranging from brute-force jamming of the spectrum to more advanced jamming attacks [2], disassociation attacks [3], and Sybil attacks [4]. Impersonation attacks, where an attacker masquerades as a legitimate transmitter, constitute a particularly significant threat since they often act as a prerequisite for deeper attacks. This raises the important question of how to secure future wireless systems against impersonation-based external threats.

Authentication and intrusion detection systems form the first line of defense against the active attack strategies. Such systems are designed to maintain message integrity, i.e., make sure that received information is originating from a legitimate source and that it has not been altered during transmission. Traditionally, authentication and intrusion detection has been implemented at higher layers (i.e., in the presentation- or application-layer of the OSI model). Today, however, it is commonly believed that the traditional authentication techniques based on cryptography are obsolete and suboptimal in some of the new use-cases of wireless communication [5]. For example, symmetric cryptography requires significant communication overhead for key-agreement and computational resources for encryption, decryption, and authentication. In machine-type communication networks, consisting of large numbers of low-power sensors and actuators that require low-latency connections, this overhead can grow beyond acceptable limits. Moreover, the low-power devices that are expected to last for decades without changing batteries might not have the computational resources required for intense cryptographic computations. New alternative security techniques are therefore researched and developed with the aim of partly replacing or complementing existing cryptographic approaches.

Recent attempts to develop such techniques can be found within the area of physical layer security [5]. Physical layer security refers to techniques that exploit properties of the physical (PHY) layer of a communication system to design schemes for secure communications. In general, physical layer security methods are designed on top of existing PHY layer protocols and exploit modulation schemes, hardware impairments, or randomness of the communication channels to secure a link (e.g., for key-agreement, encryption, jamming resilience, or authentication). These methods are promising for resource constrained communications scenarios since they make use of existing PHY layer signaling, and thus, require little or no additional overhead. Moreover, as opposed to cryptographic schemes that can be subject to brute-force attacks given enough computational resources, physical layer security schemes are considered harder to break since they rely on characteristics of the physical layer that, if properly designed, are not easily observable for an attacker.

One particular technique within the area of physical layer security is known as

feature-based physical layer authentication (PLA). Feature-based PLA schemes exploit features (i.e., signal characteristics) at the PHY layer for transmitter authentication, i.e., a receiver uses known device- or location-specific signal features in order to verify that a message is originating from an authorized source. These methods can be viewed as feature-based intrusion detection at the PHY layer that detects and filters out attacks where an attacker masquerades as a legitimate user or device to gain authorized privileges in the network. Examples of hardware-specific features that can be used for distinguishing different devices are clock skews from timing differences in the digital circuits, carrier frequency offsets and transient characteristics of received analog signals. Examples of location-specific features are frequency responses, impulse responses, received signal strengths, and angle-of-signal-arrivals.

1.1 Motivation of Thesis

Physical layer security has been proposed as a means for enhanced security in the context of ultra-reliable low-latency communication (URLLC) scenarios [6]. In particular, due to the low overhead and fast authentication at the PHY layer, feature-based PLA schemes have lately been proposed as an alternative method for authentication in mission-critical communications [7–9]. In these types of applications, we can expect to see simultaneous requirements on both system performance (i.e., reliability and latency) as well as system security (i.e., integrity, confidentiality, and availability), which will be interconnected in complex tradeoffs. Therefore, the use of feature-based PLA in such contexts will require two types of guarantees: (i) system-level delay-performance guarantees and (ii) detection performance guarantees, and analytical tools for understanding the interconnected tradeoffs.

With respect to the first type of guarantee, note that PLA often has been argued to be practical for low-latency communication scenarios. Despite this, no previous research on PLA have particularly addressed the delay impacts that would arise due to the inevitable classification errors during the feature-based authentication. False alarms, i.e., mistakenly rejecting legitimate messages, will cause packet drops that influence the delay performance. Moreover, missed detections, i.e., messages mistakenly accepted from an attacker, opens up for further attacks that can compromise the reliability and latency at a system level. Such system-level impacts need to be subject to performance bounds (i.e., guaranteed below certain levels both under normal system operation and regardless of the attackers behaviour), which is the first problem adressed in this thesis.

The second type of guarantees require analysis of optimal attack strategies, i.e., transmission strategies aimed at optimally impersonating the features of the legitimate transmitter, to obtain the worst-case detection performance. Most previous studies of feature-based PLA, with some exceptions, base detection performance evaluations on attackers that conform to the typical transmitter behavior (e.g., transmitter architecture, modulation schemes, and transmit power). However, with custom transmit hardware and software made available by relatively

cheap software defined radios, it becomes increasingly important to consider more sophisticated attack strategies against these authentication schemes. Worst-case bounds on the detection capabilities become particularly relevant for the mission-critical and URLLC scenarios where we not only want to provide strong theoretical guarantees on latency and reliability, but also on security performance.

In addition to the required performance guarantees, PLA schemes need to be better integrated into the next-generation system models, taking relevant system topologies, protocol aspects, and delay impacts into account. With increasing use of multiple antenna architectures for communication techniques like multiple-input multiple-output (MIMO) and coordinated multi-point (CoMP), the integration of PLA schemes in such systems needs to be better understood. For example, while PLA in MIMO systems has been previously studied, none of the previous works have considered PLA in distributed multiple antenna settings. Moreover, the system-level impacts beyond detection errors, like for instance impacts at the MAC layer, are often neglected.

1.2 Contributions

In this thesis, we derive mathematical tools for providing performance guarantees for feature-based PLA in mission-critical contexts. The thesis is presented in the form of six publications, referred to as Paper A-F, with individual contributions to be summarized in Section 1.3. The contributions of the thesis are summarized in the two following sections:

1.2.1 Delay Performance Guarantees

We provide mathematical methods for bounding the delay performance in wireless systems employing feature-based PLA for attack detection. These methods are developed in Paper A, B, and E. In particular, we provide:

- A queueing modeling framework that incorporates authentication-induced delays due to false alarms, missed detections, and active impersonation attacks at the medium access control (MAC) layer. The considered attack models are data injection attacks, Sybil attacks, and disassociation attacks.
- Delay performance bounds of PLA-related delays in terms of upper bounds on the delay violation probability, derived using tools from stochastic network calculus.
- An extension of the framework that encompasses distributed PLA schemes based on feature observations from multiple remote radio-heads. The model includes varying degrees of distributed processing, ranging from completely centralized (i.e., with decoding and authentication processed in a centralized fashion) to completely de-centralized (i.e., with local binary authentication and decoding decisions).

1.2.2 Detection Performance Guarantees

We provide worst-case detection performance bounds for multiple-antenna PLA schemes subject to optimal attack strategies. These results are provided in Paper C, D and F. In summary, the contributions with respect to worst-case PLA detection performance are:

- We provide the missed detection probability for a distributed PLA scheme based on channel observations at multiple distributed radio-heads.
- We derive the worst-case single-antenna attack (i.e., optimal transmit power and phase) against a particular receiver deployment (distributed in general) and an accurate approximation of the corresponding missed detection probability.
- Considering the worst-case single-antenna attacker, we provide a heuristic method for finding the optimal spatial attack position. Consequently, the corresponding missed detection probability serves as a total worst-case bound on the detection performance for the given deployment.
- We extend the analysis to an optimal multiple-antenna attacker and derive the worst-case pre-coding strategies. This analysis additionally considers attacker constraints like an ill-conditioned channel matrix, attack power constraints, and insufficient channel knowledge at the attacker.

1.3 Outline of Thesis and Included Papers

This rest of this thesis is organized as follows: Chapter 2 provides an introduction to physical layer authentication and its applications in mission-critical communication systems. It covers previous works on the topic, discusses practical integration into next-generation systems, and highlights open problems in the existing research. Chapter 3 introduces preliminary mathematical concepts that are used throughout Papers A-F. Chapter 3 covers (i) hypothesis testing, which is the basis for the physical layer authentication schemes considered in this thesis, (ii) some methods for analyzing distributions of complex Gaussian quadratic forms, and (iii) stochastic network calculus, which is the queueing analysis framework used for the delay performance evaluations. The intention behind Chapter 3 is to provide a more complete tutorial on these tools than what was allowed in the appended publications. The rest of the thesis consists of the collection of Papers A-F, each summarized in the following section:

Included Papers

The six publications Paper A-F are collaborative works of the thesis author together with the respective co-authors. The thesis author contributed with development of

Table 1.1: Summary of system assumptions and contributions of Papers A-F.

	# Antennas		Deployment		Bounds		MAC Attacks		PHY Attacks	
	$N_{\text{Rx}} = 1$	$N_{\text{Rx}} > 1$	Co-located	Distributed	Detection	Delay	Sybil	Disassociation	Power Manipulation	Spatial Position
Paper A [10]	x					x				
Paper B [11]		x	x			x	x	x		
Paper C [12]		x		x		x				
Paper D [13]		x		x	x				x	x
Paper E [14]		x		x		x	x		x	
Paper F [15]		x	x	x	x				x[1]	

concepts and theoretical results, implementation of simulation code, evaluation of numerical results, and manuscript writing. Paper A, B, and C are peer-reviewed and published papers while Paper D, E, and F are currently under submission. In the following section, the individual contribution of each paper is summarized. An overview of the considered system setup and contributions of each paper can be seen in Table 1.1.

Paper A: "On the Impact of Feature-Based Physical Layer Authentication on Network Delay Performance"

- Authors: H. Forssell, R. Thobaben, H. Al-Zubaidy, and J. Gross
- Published: *Proc. of IEEE Global Communications Conference*, Dec 2017, pp. 1–6.

This paper analyses the delay performance impacts of feature-based PLA for a single-antenna receiver. The PLA scheme is based on the complex channel gain of a line-of-sight wireless link and the attacker is assumed to be transmitting from a non line-of-sight location (e.g., from outside a factory hall). This is the first paper that

[1]Note that Paper F considers multiple attack antennas and that the power manipulation attack is equivalent to MIMO pre-coding.

uses queueing theory based on the stochastic network calculus framework to analyze the delay impacts of PLA. We provide bounds on the delay violation probability which are validated by numerical simulations. This work concluded that PLA can provide simultaneous security and low latency under strong line-of-sight channel conditions.

Paper B: "Physical Layer Authentication in Mission-Critical MTC Networks: A Security and Delay Performance Analysis"

- Authors: H. Forssell, R. Thobaben, H. Al-Zubaidy, and J. Gross
- Published: *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 4, pp. 795–808, April 2019.

This paper considers the detection and delay performance impacts of feature-based PLA in a multiple-antenna receiver. Firstly, we extend the results from Paper A to a co-located multiple-antenna receiver. In the considered multiple-antenna line-of-sight scenario, the authenticated feature is a function of the distance and angle-of-arrival with respect to the receive array. The analysis provided in this paper allows us to bound the delay impacts of PLA, in the multiple-antenna setup, using tools from stochastic network calculus. Secondly, this paper provides the models necessary to analyze the delay impacts of the active impersonation attacks known as Sybil and disassociation attacks. The main conclusion of this paper is that multiple-antenna PLA can keep the necessary latency guarantees, even under the considered impersonation attacks, for a fixed latency cost due to authentication false alarms.

Paper C: "Performance Analysis of Distributed SIMO Physical Layer Authentication"

- Authors: H. Forssell, R. Thobaben, and J. Gross
- Published: *Proc. of IEEE International Conference on Communications*, May 2019, pp. 1–6.

The main contribution of this paper is the analysis of the detection performance of feature-based PLA in a *distributed antenna system* consisting of multiple distributed radio-heads. The underlying problem motivating this setup is that PLA of a channel with respect to a single array is sensitive to impersonators transmitting from a similar angle-of-arrival. However, the distributed system model complicates the analytical derivation of the missed detection probability, which is a problem this paper provides a solution for in terms of a series expansion. This paper also provides an initial delay analysis of the distributed PLA scheme for centralized processing and with inactive attacker, which is further extended to other cases in Paper E. The main finding is that the detection performance of PLA is improved by

distributing antennas i.e., the distributed scheme provides lower missed detection probability than the co-located scenario with the same number of antennas.

Paper D: "Worst-Case Detection Performance for Distributed SIMO Physical Layer Authentication"

- Authors: H. Forssell, and R. Thobaben
- Submitted to: *IEEE Transactions on Communications*, Oct 2020.

This paper provides worst-case bounds for the detection performance of feature-based PLA under optimal single-antenna attacks. The bounds apply for PLA based on either co-located receive array or a distributed multiple array deployment. We consider two combinable attack strategies: (i) a PHY-layer attack where the attacker adapts power and phase to optimally impersonate a legitimate transmitter; and (ii) a position attack where the attacker chooses the optimal spatial position with respect to a given distributed deployment. We provide the missed detection probability under the optimal power manipulation attack and a heuristic algorithm for finding the optimal attack position. Combining these results, we are able to derive the worst-case missed detection probability for a given multiple-array deployment.

Paper E: "Delay Performance of Distributed Physical Layer Authentication Under Sybil Attacks"

- Authors: H. Forssell, and R. Thobaben
- Submitted to: *IEEE International Conference on Communications*, Oct 2020.

In this paper, we study the delay performance impacts of PLA in a distributed multiple-array scenario. This completes the delay analysis that was initiated in Paper C. The considered PLA scheme incorporates varying degrees of distributed processing, ranging from a centralized approach where authentication and decoding is performed at the centralized baseband unit, to a decentralized scenario where each remote radio-head performs independent decisions. One of the central questions is under which circumstances distributing antenna arrays is beneficial from an authentication-delay perspective. Our results indicate that the distributed approach is beneficial in terms of resilience to Sybil attacks, even under the decentralized decision scenario.

Paper F: "Worst-Case Detection Performance of Physical Layer Authentication Under Optimal MIMO Attacks"

- Authors: H. Forssell, and R. Thobaben

- Submitted to: *IEEE International Conference on Communications*, Oct 2020.

In this paper, we analyze the worst-case detection performance under optimal multiple-antenna attacks where the attacker is using MIMO pre-coding with the objective of maximizing the missed detection probability. We solve the optimal attack strategy problem under perfect channel-state information (CSI) at the attacker, imperfect CSI at the attacker, and for a power constrained attacker. Additionally, as a counter strategy, we propose to reserve a subset of silent receive antennas for reception only, in order to limit the CSI that an attacker can extract from overhearing downlink transmissions. Then, we evaluate the performance under the attack- and counter-strategies, both analytically and for recorded real-world channel traces, and show that the worst-case performance is determined by the feature-energy outside the attacker's channel range and the attack-power constraints.

1.4 Conclusions

The performance bounds derived in this thesis provide insights into the system configurations and channel conditions under which channel-based PLA is a viable option for mission-critical contexts. Some of the relevant characteristics, both with respect to delay and detection performance, are number of receive antennas, distributed vs. centralized processing, line-of-sight signal strength, channel-state information availability/knowledge, and antenna correlation. Moreover, the achievable performance depends on the attacker's capabilities in terms of CSI knowledge, power limitation, and attack position. Some of these characteristics can be influenced through system design[2] (i.e., designing antenna deployments and protocols for optimized PLA security/delay performance), while others are factors inflicted by the wireless environment. The most important conclusions of this thesis are summarized in following.

Delay Performance Impacts From the delay performance bounds in Paper A and B, first and foremost, we establish that the considered channel-based PLA schemes can be deployed while maintaining latency requirements on a mission-critical level. These results were however contingent on significant line-of-sight paths from legitimate transmitters to the PLA receiver. Based on that observation, we conclude that the considered PLA approach (i.e., based on line-of-sight received power and angle-of-arrival) would be relevant in fixed indoor deployments like in industrial factory automation. While device mobility was not considered in these papers, we discuss how the results can be extended to mobile scenarios, something that would make the results valuable for vehicle-to-roadside communications where line-of-sight channels are often assumed.

[2]In addition to this section and the included papers, such design choices are discussed in Section 2.3.

PLA With Distributed Antennas With respect to the distributed receiver architecture, considered in Paper C, D and E, our results show significant benefits in terms of worst-case PLA performance. Conceptually, the explanation is that the angle-of-arrival and received power patterns with respect to multiple receivers become increasingly difficult for the adversary to impersonate. The main conclusion is that the worst-case receiver operating characteristic is improved (i.e., a lower missed detection probability for a given false alarm probability can be achieved) by distributing antennas to multiple locations. In the mission-critical contexts considered in this thesis, such performance benefits can practically be traded against either improved latency performance for a given security level, or vice versa.

Attack and Counter Strategies In Paper D, we find that the optimal attack strategy against a single-array receiver is to choose the same (or mirroring) angle-of-arrival and adapt the power to match the legitimate device. This type of attack has large impact on the detection performance, a problem which has three potential solutions: (i) Secure physical exclusion regions so that the attacker cannot transmit from such favorable locations, (ii) use PLA based on multiple distributed receivers (as argued in the previous paragraph), or (iii) consider extended PLA schemes based on combinations with other PHY-layer features. In Paper F, concerning the multiple-antenna attacker strategies, we identify channel characteristics (e.g., the feature energy outside the range of the attackers channel matrix) that are determining factors for the attack success probability. Such observations can be practically used in real systems by analyzing channel measurements to identify critical attack positions. Moreover, we conclude that the proposed counter strategy can improve the worst-case detection performance with 1-2 orders of magnitude in the considered scenario.

Future Work

Worst-case performance analysis constitutes a key component for the practical integration of channel-based PLA in real-world systems. Towards that goal, we have identified several items that would be interesting to adress in future research:

- As discussed in Chapter 2, artificial intelligence and machine learning can also be used for feature-based PLA. In previous works, no closed-form detection performance for such schemes has been derived, and we anticipate that general performance bounds like the ones provided in this thesis are challenging to obtain. However, a problem that can be adressed is empirical comparisons between performance of machine learning PLA (like e.g., provided in [16]) and the statistical hypothesis testing setup and bounds obtained in this thesis.
- Experimental evaluation based on software defined radios. In particular, it would be relevant to implement the PHY-layer attack strategies, record empirical channel distributions, and evaluate the attacker impacts compared to the derived performance bounds.

- Transmitter mobility and tracking of channel-state information are open theoretical extensions to the results derived in this thesis that would broaden the applicability in real-world systems. A potential scenario to consider would be angle-of-arrival-based PLA of vehicular communications (e.g., vehicle-to-roadside or from unmanned aerial vehicles) communication.

- Cross-layer intrusion detection in wireless control systems is one way to integrate channel-based PLA in larger intrusion or fault detection contexts. Cross layer schemes could benefit from exploiting correlations between sensing and actuation information and wireless features at the PHY-layer. For instance, this idea connects well with the concept of real-time situational awareness in cyber-physical systems [17], where timely and accurate incident response is the target.

1.5 Other Contributions

In addition to Paper A-F that comprise this thesis, the author has during this period also been contributing to the following research works:

Paper G: "Feature-Based Multi-User Authentication for Parallel Uplink Transmissions"

- Authors: H. Forssell, R. Thobaben, J. Gross, and M. Skoglund

- Published: *Proc. of 9th International Symposium on Turbo Codes and Iterative Information Processing*, Sep. 2016, pp. 355–359.

In this paper, we provide a factor-graph framework for PLA of multi-user uplink transmissions over time-variant channels. Through this approach, we derive the closed-form a posteriori attack probability that can be used as soft information at the PLA receiver. These results show how the receiver can exploit the multi-user setup, by using the cross-channel correlation of the large-scale fading parameters, for enhanced PLA performance.

As opposed to Paper A-F, this paper targets protocol and PLA scheme design, and does not consider worst-case performance analysis, and for this reason, the paper was excluded from the thesis.

Paper H: "A Novel Low-Complexity Power-Allocation Algorithm for Multi-Tone Signals for Wireless Power Transfer"

- Authors: B. A. Mouris, H. Forssell, and R. Thobaben

- Published: *Proc. of IEEE Wireless Communications and Networking Conference*, Seoul, Korea (South), 2020, pp. 1-6.

This paper proposes a novel low-complexity algorithm for allocating power to multi-tone signals for wireless power transfer. The algorithm, referred to as truncated maximum-ratio transmission (TMRT), performs maximum ratio transmission power allocation on the subset of the strongest channels. Simulation results confirm that the proposed TMRT algorithm achieves a performance very close to the optimal power allocation, despite its very low complexity, and significantly outperforms other low-complexity solutions.

In this work, the thesis author contributed with collecting multi-carrier channel measurements using a universal software radio peripheral (USRP) radio platform. The channel measurements were used for numerical results that validate the effectiveness of the algorithm.

CERCES Project: "Testbed Demonstrator" The research for this thesis was conducted as a part of the CERCES[3] project. Within the CERCES project, a testbed was developed with the purpose of demonstrating new security techniques in critical cyber-physical systems. A PLA scheme, annotating packets as legitimate or suspicious based on PHY layer features, was implemented at the PHY layer of IEEE 802.11g running on a USRP radio platform. This PLA scheme was in the testbed integrated in a real-time intrusion detection system running on a remote controlled lego robot.

[3]Center for resilient critical infrastructures.
`https://www.kth.se/dcs/research/secure-control-systems/cerces/`

Chapter 2

Physical Layer Authentication in Mission-Critical Communications

This chapter provides an overview of the background concepts related to the integration of physical layer authentication (PLA) in mission-critical communications. We outline the challenges associated with mission-critical communication scenarios, introduce the concept of PLA, and survey previous work on the subject. In the final section, we discuss design and deployment aspects for integrating PLA in practical systems.

2.1 Challenges in Mission-Critical Communications

As of today, fifth generation (5G) mobile networks are being commercialized, and sixth generation (6G) applications and technologies are at an early stage of research. These systems are often conceptualized in terms of the three areas (i) enhanced mobile broadband (eMBB), (ii) massive machine-type communication (mMTC), and ultra-reliable low-latency communication (URLLC), where each area is composed of current and envisioned applications as well as its own range of challenges. The area of URLLC is arguably the most challenging due to its conflicting goals of realizing very low latencies with simultaneous very high requirements on reliability. URLLC is anticipated to provide the communication performance required for mission-critical applications like automated industries (i.e., Industry 4.0), remote surgery, autonomous driving, unmanned arial vehicles (UAV), smart metering, and surveillance. With URLLC, new constraints emerge both in terms of communication and security requirements.

2.1.1 Communication Requirements

Typically, the low-latency requirements for mission-critical applications are in the order of less than 1 ms latency with higher than 99.999% reliability [18]. Generally

speaking, enhanced reliability of wireless communications is achieved through redundancy and diversity, realized in either time (e.g., re-transmissions), frequency, or space (multiple-antenna diversity). Such methods come with costs in terms of additional transmission time, processing time, or signaling overhead. Therefore, under the complexity constraints of the envisioned applications, latency and reliability are conflicting requirements which is the central challenging tradeoff in URLLC system design. The latency sensitive information in autonomous control applications often consists of small transmission payloads, as opposed to the more data-driven human-centered applications. This poses another challenge, since the traditional tools for throughput, scheduling, and security analysis do not apply straightforwardly to URLLC scenarios [19].

Enabling technologies for solving the low-latency high-reliability tradeoff relates to the development of new system architectures and customized PHY and MAC-layer designs. Reduced transmission-time intervals, non-orthogonal multiple-access, device-to-device communication, and frequency hopping are some of the discussed potential enablers [20]. The diversity of multiple-antenna systems is furthermore a potential solution for providing the mission-critical reliability and availability levels. For instance, massive and distributed MIMO systems are considered for the realization of URLLC [21].

2.1.2 Security Challenges

Many of the considered use-cases are associated with automation of societal functions like traffic, manufacturing, industry, and the security and safety of these applications is ultimately important for protecting human lives and economic values. However, paired with the above-mentioned communication requirements, providing security in these application poses additional challenges.

Overhead/Processing Limitations Firstly, cryptographic security schemes are often not suitable for mission-critical and URLLC scenarios due to multiple reasons. One reason is that the required encryption and decryption algorithms may be computationally too complex to realize the strict latency requirements [6]. Moreover, cryptographic schemes require transmission overhead in terms of dedicated signaling for key-agreement, introducing additional delays that might not be tolerable in low-latency applications. Thirdly, due to the often very small data payloads in mission-critical application, the actual transmission overhead for authentication can begin to constitute a significant part of the entire message. As a concrete example based on IEEE 802.15.4 illustrates [8], with a small payload of 32 bytes the overhead for AES-128 CMAC encryption already constitutes a 20% overhead.

Reliability/Availability vs. Security Another perspective on the challenge of secure mission-critical communication is obtained by comparing it to the more traditional human-centered scenarios (i.e., eMBB). In human-oriented communication,

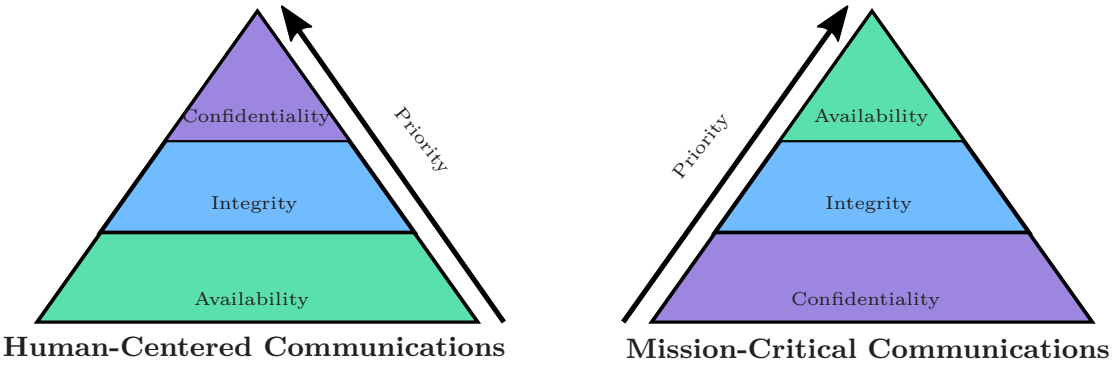

Figure 2.1: Shift of priorities between human-centered and mission-critical communication scenarios.

data confidentiality followed by integrity form the central priorities while service availability and security overhead typically have lower priority. The reason is that the applications (i.e., mobile broadband, e-mail, video- and audio-streaming) are not as time-critical and reliability issues can be solved by re-transmissions. However, as illustrated in Figure 2.1, in mission-critical communications the order of concern is reversed [22]. Service availability is of highest priority in mission-critical scenarios since the applications are typically supposed to run uninterrupted over long time spans. Service outages can have severe consequences both in terms of human safety and economic costs. The second highest priority is message integrity. For example, it is of vital importance that sensor and actuation information in a closed-loop control application is not altered during transmission, and thus, it must be assured that the received data indeed stems from the claiming source. Finally, confidentiality is of lowest priority as in automation applications the reading of sensor and actuation information poses a comparably smaller threat to the controlled plant[1].

CSI Acquisition Physical layer security is considered a potential solution to some of the security related challenges of mission-critical communications. Physical layer security refers to techniques that exploit properties of the physical (PHY) layer of a communication system to design schemes for secure communications. In particular, channel-state information (CSI) is often used as a source of randomness for physical layer security. However, in URLLC and mission-critical communications, due to the strict latency requirements, it is not always possible to obtain full CSI for each communication time-slot [6]. Such limitations constitute a challenge for many physical layer security methods since they often rely on updated and accurate CSI. For general physical layer security, this challenge depends on to which extent the CSI is available to both transmitter and receiver. However, for CSI-

[1]This is not to say that information confidentiality is insignificant in mission-critical contexts, but rather an argument for why the order of priority is shifted.

based physical layer authentication, where the location-specific CSI is exploited for device authentication (introduced in detail in Section 2.2), only the receiver needs to obtain the CSI. Multiple-antenna systems and location-based beam-forming can potentially provide a solution to the CSI-acquisition for physical layer security in URLLC scenarios [6]. For such methods, only legitimate transmitter locations are relevant for the receiver; however, note that a line-of-sight or dominating reflecting path from transmitter to receiver is required.

Attacks in the Wireless Domain Impersonation-based attacks pose a significant threat against future wireless mission-critical communication systems. In the wireless domain, impersonation of another transmitter gives the attacker legitimate privileges which opens up the possibility for other types of attacks like denial-of-service (DoS), disassociation attacks [3], and Sybil attacks [4]. Apart from impersonation-based vulnerabilities, there are other types of attacks like eavesdropping and jamming attacks [2]. In Table 2.1, we summarize and define a range of wireless domain threats against mission-critical communication systems. The first types of attacks (marked in blue) are impersonation-based and can be detected by PLA, while the latter types (marked in red) require other types of security techniques.

PLA is one potential solution for the described security challenges within mission-critical communications. In the next section, we introduce the concept of PLA and previously proposed PLA schemes.

2.2 Physical Layer Authentication

Authentication is the process by which a receiver verifies that a message is originating from an authorized source. PLA schemes exploit properties of the PHY layer for fast authentication prior to passing messages to higher layers. This section introduces the concept of PLA with a detailed description of feature-based schemes in Section 2.2.1, an overview of tag-based approaches in Section 2.2.2, and a survey of previous work in Section 2.2.3.

2.2.1 Feature-Based Physical Layer Authentication

In feature-based PLA schemes, authentication is achieved based on extracting physical layer characteristics that in some way uniquely identify the transmitter. The authentication task is formulated as a classification problem, aimed at determining whether the observed feature is consistent with some prior knowledge on the feature of the legitimate transmitter. In general, this classification can be done through either statistical hypothesis testing or machine learning techniques. Since the PLA schemes considered in this thesis are based on the hypothesis testing formulation, we will begin by illustrating a generic setup for these types of schemes.

Table 2.1: Summary and descriptions of impersonation-based attacks in wireless systems.

Attack Type	Description
Impersonation	Spoofing the identity of a legitimate entity in the network. For example, faking the cryptographic credentials of an authentic user to get access to protected functions in the network.
Data Injection	Injecting malicious transmissions with false information into the network. For instance, in a cyber physical system, an adversary can attempt to send fake sensor or control messages with the aim of driving the system into an unsafe state.
Denial-of-Service (DoS)	Reducing or eliminating the service (i.e., communication resources in a wireless system) available for normal use of the system. This can be called a distributed DoS (DDoS) attack when launched from multiple nodes that are under control by the attacker.
Disassociation	In wireless systems where devices can connect/disconnect from the network, an attacker can exploit this function to disconnect legitimate devices and interrupt their service.
Sybil Attack	An attacker impersonates multiple fake identities which is interpreted as a large number of devices that request network services. This is a form of DoS attack since the service to legitimate devices will be reduced.
Jamming	Interrupting communication between legitimate parties by transmitting interfering signals. Jamming can either be achieved in a 'brute force' manner; the jammer transmits noise with high power in the frequency band of the legitimate communication, or by more intelligent means; the jammer injects signals which are designed to maximally interfere with the targeted communication link.
Eavesdropping	Receiving and decoding transmissions. In mission-critical systems, this can be used to listen and learn system characteristics (e.g., dynamics of a physical control-loop) for launching optimally designed active attacks.

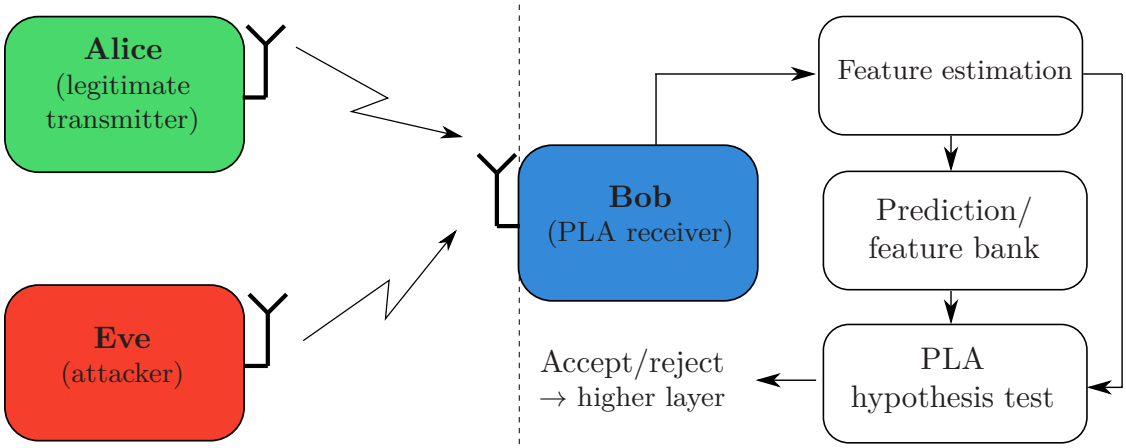

Figure 2.2: Generic model for feature-based physical layer authentication of wireless transmissions.

The generic model for a feature-based PLA scheme is depicted in Figure 2.2. The setup consists of a legitimate transmitter Alice, a legitimate receiver Bob, and an attacker Eve that attempts to transmit messages to Bob while claiming to be Alice. Assuming that messages are received at discrete time-slots indexed by k, Bob's objective is to determine if the transmission at time k is originating from Alice or Eve. In order to achieve this, Bob utilizes that signals transmitted from a specific source can be characterized by a transmitter-specific feature vector, and estimates the feature vector $\hat{\mathbf{x}}(k)$ for the received signal, which is modeled according to

$$\hat{\mathbf{x}}(k) = \begin{cases} \mathbf{x}_A(k) + \mathbf{w}(k) & \text{if} \quad \mathcal{H}_0 \\ \mathbf{x}_E(k) + \mathbf{w}(k) & \text{if} \quad \mathcal{H}_1, \end{cases} \tag{2.1}$$

where $\mathbf{x}_A(k)$ and $\mathbf{x}_E(k)$ are the transmitter-specific features of Alice and Eve[2], respectively, $\mathbf{w}(k)$ is measurement noise, $\mathcal{H}_0$ represents the hypothesis that the transmission is legitimate, and $\mathcal{H}_1$ represents the hypothesis that it is an impersonation attempt from Eve. In this formulation, the feature vector $\hat{\mathbf{x}}(k)$ could represent any extracted feature of the PHY layer signals, and particular examples will be introduced and discussed in Section 2.2.3 and 2.3.1. In general, the transmitter-specific features follow a dynamical processes model

$$\begin{aligned} \mathbf{x}_A(k) &= f_A(\mathbf{x}_A(k-1)) + \mathbf{z}_A(k) \\ \mathbf{x}_E(k) &= f_E(\mathbf{x}_E(k-1)) + \mathbf{z}_E(k), \end{aligned} \tag{2.2}$$

where $f_i()$ and $\mathbf{z}_i(t)$ for $i \in \{A, E\}$ represent the dynamics of the feature vectors and process noise, respectively. A special case with time-invariant features is obtained by assuming $f_A(\mathbf{x}_A(k-1)) = \mathbf{x}_A$ and $f_E(\mathbf{x}_E(k-1)) = \mathbf{x}_E$.

[2]Note that Bob generally does not know Eve's feature vector in advance.

PLA of a transmission is performed at Bob as a binary hypothesis test

$$d(\hat{\mathbf{x}}(k), k) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} T, \tag{2.3}$$

where $d(\hat{\mathbf{x}}(k), k)$ is some appropriately derived discriminant function and T is a design threshold.

Discriminant Function Design Appropriate design of discriminant function relates to the choice of feature and assumptions on underlying statistical distributions. Moreover, as the discriminant function depends on the legitimate feature distribution, the design also relates to feature learning. Typically, the discriminant function represents some kind of distance measure with respect to the predicted legitimate distribution. For instance, under proper complex Gaussian feature vector, the discriminant function $d(\hat{\mathbf{x}}(k), k)$ is obtained by deriving the log-likelihood of the observation $\hat{\mathbf{x}}(k)$ conditioned on the legitimate hypothesis $\mathcal{H}_0$. This results in the generalized likelihood-ratio test (GLRT) setup where the discriminant function is given by the Mahalanobis distance. Particular discriminant function design under Neyman-Pearson and GLRT setups will be provided in Chapter 3.1 as well as in the collection of Papers A-F. Although in this section we consider the time-variant case to maintain generality, note that throughout the appended papers we assume time-invariant feature statistics, Moreover, we note that the problem of feature learning is outside the scope of this thesis; however, some aspects related to system-level integration of PLA are discussed in Section 2.3.2.

Error Probabilities As feature-based PLA is based on hypothesis testing, two types of errors can occur: (i) A false alarm (also referred to as a Type-I error) where a legitimate message is rejected and (ii) a missed detection (also referred to as a Type-II error) where an illegitimate message is accepted. The probabilities of these events are used as performance metrics for a given PLA scheme. Mathematically, the probability of false alarm is defined as

$$p_{\mathrm{FA}}(T) = \mathbb{P}\left(d(\hat{\mathbf{x}}(k), k) > T | \mathcal{H}_0\right), \tag{2.4}$$

and the probability of missed detection as

$$p_{\mathrm{MD}}(T) = \mathbb{P}\left(d(\hat{\mathbf{x}}(k), k) < T | \mathcal{H}_1\right). \tag{2.5}$$

From these definitions, we can observe that $p_{\mathrm{FA}}(T)$ is a decreasing function with T, while $p_{\mathrm{MD}}(T)$ is an increasing function with T, and that the choice of threshold T results in a tradeoff between false alarms and missed detections. In practice, given that the legitimate feature distribution is known, Bob will compute the threshold from (2.4) for an acceptable false alarm probability p_{FA}.

2.2.2 Tag-Based Physical Layer Authentication

In addition to the feature-based PLA schemes, there exists tag-based PLA approaches that are based on a secret modulation signal superimposed to the transmitted signal and acts as a signal watermark [23]. Essentially, these schemes are similar to feature-based schemes, except for that the transmitter-specific feature, which needs to be known by both parties, is deliberately injected at the transmitter side.

Tag-based approaches come with both benefits and drawbacks compared to feature-based PLA. One the one hand, tag-based PLA can be considered more robust since the feature-based approach relies on randomly varying features that cannot be directly influenced by the designer. One drawback, on the other hand, is that legitimate tag-sequences can leak to an eavesdropper, which is why covertness is a central design-problem in tag-based schemes. Moreover, the injected tags require additional resources in terms of bandwidth, transmit energy, and processing resources, which comes at a cost of reduced decoding performance and overhead. Finally, since pre-agreed secrets are required for choosing the secret modulation patterns, the problem of overhead for key-agreement cannot be directly solved by tag-based schemes. For these reasons, although tag-based schemes could be useful in certain scenarios, they appear less promising for meeting the mission-critical requirements considered in this thesis.

2.2.3 Previous Work

In this section, we survey previous work on PLA and previously considered feature choices. Many of these works are also discussed in the surveys [24, 25] and overviews provided by [5, 26].

Hardware-Based PLA Hardware-based PLA schemes exploit slight differences between hardware chipsets to identify/authenticate users or devices. PLA schemes based on the carrier frequency offset (CFO) has been proposed in [27, 28]. Carrier frequency offsets represent the hardware imperfections of the local oscillators that vary randomly across different transmitters. Due to the time-variant nature of the CFO, [27] combines feature tracking based on Kalman filters with binary hypothesis testing. In [29], PLA is studied based on the device-specific unintentional RF emissions which are classified based on statistical properties. In [30], PLA based on the device-specific IQ imbalance is studied in the context of amplify-and-forward systems. In [31], a model-based approach is developed for PLA based on RF-chain imperfections. Moreover, improvements of using multiple-antenna receivers for PLA based on radiometric-features is studied in [32].

RSSI-Based PLA Using the received signal strength indicator (RSSI) for detecting node replacement attacks was considered in [33]. In [34], a more resilient feature, consisting of the RSSI profile measured at multiple neighbouring wireless

receivers, is used for detection of identity-based attacks in wireless sensor networks. In [35], an energy-ratio detector is used to detect pilot spoofing attacks based on the received power. In [36], RSSIs and inter-vehicle distances are used within a Kalman-tracking framework for authentication of vehicle to everything (V2X) communications. A handover authentication mechanism jointly based on RSSI and packet error rates is proposed in [37]. Moreover, in [38], a PLA scheme based on RSSI is developed for detecting identity-based attacks and is shown to be able to localize the attacker device.

Channel-Based PLA Using Hypothesis Testing Feature-based PLA based on the channel frequency response was first proposed within a hypothesis testing framework in [39]. This concept was then extended to time-variant channel [40–42], MIMO channel [43], for detection of Sybil attacks [44], to a sensor network context in [45], in a reinforcement learning context in [46], and in a game-theoretic framework in [47]. In [48], a PLA scheme based on the time varying channel impulse response (CIR) is proposed in which the hypothesis test threshold is set adaptively based on the receiver SNR. Additionally, their work is extended with a quantization algorithm for the CIR-based PLA scheme in [49, 50]. In [51], a PLA scheme based on the complex channel matrix in a MIMO system is proposed. A similar MIMO setup is considered in [52]. In [53], the complex channel gains from multiple sensors to a centralized anchor node are used for detection of spoofing attacks in an IoT network. The authors of [54] propose a cross-layer authentication method including PLA hypothesis tests for local security in IoT networks. In [55], PLA based on the channel-state information is considered for distributed ad-hoc sensor networks. The work in [56] proposes a challenge-response authentication mechanism based on the OFDM channel which is also extended to relay communications in [57]. A similar challenge-response PLA scheme is investigated in [58]. Hypothesis-testing PLA schemes based on the observed power spectral density, closely related to the frequency response, was proposed in [59], and based on multi-carrier CSI in [60].

Angle-of-arrival (AoA) based PLA AoA based schemes have mainly been considered for vehicle-to-roadside communications. In [61], both PLA and physical-layer secret key-agreement is considered based on estimated AoA from vehicles transmitting with line-of-sight and a known GPS positions. A similar AoA-based scheme is studied in [62] and the results are validated by measurements from a software-defined radio platform.

Machine Learning-Based PLA In contrast to the hypothesis testing-based approaches, there are several previous works that investigate the use of machine learning for the transmitter classification problem. A machine learning approach is taken in [63] where the proposed algorithm tracks the time variant channel frequency response, modeled as a Gaussian process. In [64], the authors propose a deep-learning based approach for authentication of IoT devices. PLA methods

based on support vector machines and linear Fisher discriminant analysis based on multiple physical layer features are proposed in [65]. In [66], a machine learning approach for threshold-free channel-based PLA is proposed and validated on a USRP platform. In [16], a comparison between statistical and machine learning techniques for PLA is provided.

Adversarial Strategies For more resilient performance evaluation of feature-based PLA, one can consider attackers that are aware of the employed PLA scheme and use various attack strategies. Such strategies consist of eavesdropping for legitimate feature information and transmission strategies aimed at maximizing the probability of attack success. There are fairly straightforward attack strategies against PLA based on single dimensional features like CFO and RSSI. Legitimate CFOs can be impersonated by eavesdropping and adapting the transmit frequency to match the legitimate transmitter. Moreover, RSSIs can be altered by manipulating the transmit power. For PLA based on more diverse channel features, like e.g., multiple-input multiple-output (MIMO) channels, the question of optimal attack strategies becomes more involved. In [52], the performance evaluation of the multiple-antenna PLA scheme considers an attacker that uses pre-coding based on correlated observations to optimally mimic the legitimate channel feature. In a similar setup, [67] derives the outer region of the achievable detection performance for a MIMO/OFDM-based PLA scheme. Furthermore, performance evaluations in [47, 51, 68] also considers variations of PLA attack strategies.

Information Theoretic Analysis Authentication has furthermore been studied within information theory. However, most commonly based on a joint source of randomness (i.e., shared secret key or observation) [69]. The work in [70] provides information theoretic bounds for the hypothesis-testing based authentication, where bounds are provided in terms of the Kullback-Leibler divergence between the legitimate and the attacker distributions.

Multi-Feature Authentication Some previous works have considered PLA based on a combination of multiple physical layer features. For instance, [71] considers a general feature vector and hypothesis testing based on Kolmogorov-Smirnov tests. The final authentication decision is based on majority vote. Also, the work in [72] addresses the problem of a generic PLA scheme for combining multiple weighted device-specific features into a joint authentication decision. In [73], a cross-layer authentication approach is developed, partly based on the in-phase/quadrature-phase imbalance (IQI) and CFO observed at the PHY layer. In [74], the authors consider centralized channel-based PLA with observations from multiple reception point. The focus is on decision fusion based on compressed sensing. In addition, the previously mentioned works [34, 53] are multi-feature authentication schemes in the sense that they combine features from multiple reception points.

Key/Tag-Based PLA Recall from Section 2.2.2 that tag-based PLA is based on a secret modulation signal superimposed to the transmitted signal that acts as a signal watermark. Tag-based PLA was first introduced in [23]. The same authors investigate the tradeoff between security and decoding performance due to tag-power allocation in multi-carrier systems in [75]. Moreover, similar tag/watermark-based PLA methods have been proposed for detecting wormhole attacks in mobile ad-hoc networks [76], industrial IoT scenarios [77], digital television transmissions [78, 79], and wireless OFDM [80] and MIMO systems [81, 82]. In [58], devices are authenticated based on a secret phase modulation applied to a multi-carrier transmission. Finally, in [83] a key-based PLA approach is analyzed where the secret key is generated based on the physical layer channel.

PLA for Mission-Critical and URLLC Feature-based PLA has previously been proposed for security in URLLC [7, 9]. In [7], PLA is based on frequency domain CSI and a Gaussian mixture model is used for classification. The results are evaluated based on measurements from a software defined radio platform. In [9], the method is extended to a range of other supervised classification methods, including stochastic gradient descent, random forest, support vector machine, and linear discriminant analysis. These works focus on the detection performance and complexity, and argue that the methods are suitable for URLLC and mission-critical scenarios. However, as opposed to the analysis conducted in this thesis, these works do not consider the delay impacts due to erroneous authentication decisions that can be detrimental in these types of applications.

2.3 Practical Design and Deployment Aspects

Practical integration of PLA in mission-critical systems requires knowledge of how security and reliability aspects relate to higher level system design choices. For instance, what constitutes appropriate feature choices depends on the particular system architecture, channel conditions, and protocol design. Moreover, the use of PLA for enhanced security is just one component in a larger system context, which motivates why it is important to consider higher system-level performance impacts of PLA as well. With the mission-critical challenges and concept of PLA established, this section discusses such practical design and integration aspects.

2.3.1 Feature Requirements

Generally speaking, there are three categories of feature requirements for feature-based PLA schemes: (i) *predictability*, (ii) *observability and reproducibility*, and (iii) *accessability*. Here, let us pick up the notation from Section 2.2.1. Predictability quantifies to how well Bob can predict the feature state $\mathbf{x}_A(k)$ of Alice given previous estimates $\hat{\mathbf{x}}(k')$ for $k' < k$. This is related to the accuracy of Bob's estimates and the amount of correlation in the feature processes in (2.2). Observability we

define as how exposed the feature state $\mathbf{x}_A(k)$ is to the adversary Eve, i.e., Eve's ability to estimate the state of Alice's feature by eavesdropping. In a similar way, reproducibility refers to Eve's ability to, given an estimate of the legitimate state, influence her own feature to mimic Alice's. Finally, accessibility refers to the amount of processing required at Bob in order to obtain the feature estimates $\hat{\mathbf{x}}(k)$ for each transmission. The requirement for a good feature is that it is predictable and accessible for Bob and difficult or impossible to observe and reproduce for Eve.

Based on these requirements, we will now discuss the characteristics of typically employed features for PLA. The discussed features are summarized in Table 2.2. We begin with hardware-specific features (i.e., features of the transmitting chipset that are due to hardware and manufacturing imperfections).

Carrier Frequency Offset Recall that carrier frequency offsets (CFO) are a commonly proposed feature for PLA. CFOs result from the offset $\Delta f_{\mathrm{AB}}(k) = f_{\mathrm{A}}(k) - f_{\mathrm{B}}(k)$ in the local oscillators when up-converting baseband signals, where $f_{\mathrm{A}}(k)$ and $f_{\mathrm{B}}(k)$ are respectively Alice's and Bob's realizations of the targeted carrier frequency f_c. Experiments using software-defined radios have shown that estimated CFOs are sufficiently distinct across different chipsets to reliably authenticate at the PHY layer [27]. Due to temperature variations CFOs experience moderate time-variations that can be modeled by linear dynamical systems. One issue with using CFO as a feature is that Eve can estimate the offsets $\Delta f_{\mathrm{AE}}(k)$ by eavesdropping and alter her own carrier frequency to $f'_{\mathrm{E}}(k) = f_{\mathrm{E}}(k) + \Delta\hat{f}_{\mathrm{AE}}(k) = f_{\mathrm{A}}(k) + \epsilon_{\mathrm{E}}$, where ϵ_{E} is Eve's estimation error. This means that Eve, equipped with appropriate hardware that allows such carrier frequency modification, quite easily can impersonate Alice and pass the authentication test. On the other hand, a big advantage is the CFOs accessibility; since the offset needs to be compensated for coherent demodulation, CFO estimates are already estimated at the physical layer in most communication standards.

Clock Skews This feature is resulting from timing differences in the digital circuits can also be used for authentication purposes. In [84], experimental results indicate that clock skews, estimated from the IEEE 802.11 Time Synchronization Function (TSF), remain constant across several experiments and differ around 10-20 ppm across different transmitter chipsets. Given a receiver function such as above-mentioned TSF, the clock skew is accessible to Bob and also observable at Eve. However, it is difficult for Eve to reproduce the observed feature without switching chipset since clock-skew offsets generally are not controllable by the user.

Transients Transients occurring at the beginning of captured wireless signals possess chipset-specific characteristics that as well can be exploited for PLA. Effective transmitter identification has been demonstrated for radars, IEEE 802.11 and Bluetooth devices [85]. This approach however requires a feature-extraction process to reduce the large-dimensional observed transient to a feature vector. The

extracted feature vectors in [85] were observed to be predictable over time for a fixed position and antenna polarization. However, changing positions and antenna polarization degraded performance. Transients could in principle be recorded and replayed by Eve. However, since results in [85] indicated that transients were influenced by distance and antenna orientation and not only chipset, Eve would need to take these into consideration which further complicates reproducing the transients from Alice.

As opposed to the above-mentioned features, location-specific features are characteristics of the wireless communication channel that are specific to the spatial location of the transmitter. In the following, we discuss the commonly used location-specific features.

Received Signal Strength Received signal strength is a measure of received power, commonly evaluated in most wireless receivers, that depends on the distance between transmitter and receiver and can be used to differentiate between different transmitters. For instance, experiments in [33] show that received signal strength indicator (RSSI) measurements allow a receiver to efficiently resolve transmitter distances of 1.5 meters. However, the scalar RSSI values can only position its source in one dimension and are furthermore dependent on transmission power, something that makes the scheme vulnerable to feature spoofing attacks where the adversary varies transmission power to mimic the legitimate transmitter.

Channel Frequency/Impulse Response Due to multi-path propagation of radio-waves through the wireless environment, the wireless channel can provide a more diverse location-specific feature, often represented as multi-carrier frequency responses or impulse responses. The impulse response captures the multi-path environment in the time-domain and represents the signal amplitude and phases received at different delays caused by the difference in propagation distance in the specific paths. The frequency response is the frequency-domain equivalent to the impulse response and represents how the wireless channel attenuates different frequencies. Additionally, both are influenced by changes in the wireless environment (e.g., due to transceiver or environmental mobility) which cause certain paths to experience fading. Using the frequency response for location-based PLA was first proposed in [39].

Frequency responses are easily accessible in orthogonal frequency division multiplexing (OFDM)-based systems (e.g., LTE, IEEE 802.11) since they are estimated as channel-state information (CSI) for channel equalization. Similarly, impulse responses are accessible in code-division multiple access (CDMA)-based systems such as 3G. Important parameters for the predictability of the multi-path dependent features are coherence time and coherence bandwidth. Coherence time quantifies the time period over which the channel can be expected to remain constant. Hence, for the multi-path channel to be predictable at Bob, it is required that the period between transmissions is smaller than or at least in the order of the coherence

time. That is, we need relatively slowly changing wireless channels, which are typically observed for low mobility transmitters and receivers in slowly changing environments. The coherence bandwidth, which is inversely proportional to the length of the impulse response (i.e., the delay spread), quantifies the bandwidth within which the channel can be expected to remain constant, i.e., the frequency selectivity of the channel. To obtain a diverse feature that uniquely characterizes separated transmitter locations, a frequency-selective channel with small coherence bandwidth relative to the communication bandwidth is preferable. This is generally the case in communications over longer distances that generate larger delay spreads (due to a rich variety of reflection and scattering paths). This implies that frequency/impulse responses are more appropriate as features to authenticate in large scale cellular deployments than in short-range indoor systems such as Bluetooth and WiFi (except potentially for ultra-wideband systems that could provide rich frequency responses even at large coherence bandwidths).

In rich scattering environments, multi-path channels at positions separated by more than half a wavelength experience uncorrelated fading [39]. Given such assumptions, the attacker could not infer the legitimate channel realization by eavesdropping. However, the attacker could use raytracing algorithms and through knowledge of the deployment topology approximate the legitimate channel distributions.

Multiple-Antenna Channel and Angle-of-arrival (AoA) Multiple-antenna receivers can add additional dimensions to the feature vector and improve PLA performance. For instance, such receivers allow a differentiation between the angles of incoming transmissions based on the relative phases of the received signal at each antenna element. AoA profiles can be used for PLA since the AoAs will depend on the physical location of the transmitter. However, dominant line-of-sight or reflective paths from the transmitter to the receiver are required for the location-specificity of the AoA. Furthermore, a relatively stationary deployment (e.g., wireless sensor at fixed positions) or mobile devices with predictable trajectories (e.g., sensors in factory automation or vehicular networks) are required for the predictability of the AoA as a feature. Note that it is difficult for an adversary to infer the legitimate AoA by eavesdropping. However, through knowledge of the specific deployment, Eve can choose a position to obtain a similar AoA profile as Alice. For a scenario with larger distances (> 100 m) without dominating LoS path, predictability requires that changes in the environment and mobility of transmitter and receiver can be tracked. An attacker may be able to employ raytracing for estimating the AoA; however, utilizing this knowledge for impersonation in real-time is very complicated or impossible. This problem can potentially be alleviated by authenticating the AoA profile with respect to multiple distributed antenna arrays, as argued in Papers C-E.

Table 2.2: Evaluation of requirements for commonly used features.

	Predictability (at Bob)	**Accessability (at Bob)**	**Observability / Reproducibility (at Eve)**
Chipset-Specific			
Carrier frequency offset	Moderately time varying. Predictable by e.g., a Kalman filter.	Already estimated at the PHY layer.	Easily estimated and reproduced by eavesdropping Alice
Clock skew	Small time variations.	Requires no additional processing if TSF exists (as e.g., in IEEE 802.11).	Simple to observe but complex to reproduce.
Analog transient	Stable for fixed transmitter position and antenna polarization.	Requires feature extraction process.	Record-and-replay attack (requires correct distance and antenna polarization).
Location-Specific			
Received signal strength	Stable for fixed transmitter positions with 1.5 m resolution.	Estimated at PHY layer in most communication standards.	Easily observable and reproducible by varying transmit power.
Multi-carrier frequency response	Predictable within the order of the coherence time.	Directly accessible as CSI in OFDM-based systems (e.g., IEEE 802.11 and LTE).	Difficult to estimate from uncorrelated adversary channel (further than $\lambda/2$ from Alice).
Impulse reponse	_ ” _	Directly accessible as CSI in CDMA-based systems (e.g., 3G).	_ ” _
Angle-of-arrival	For LoS scenarios with fixed transmitter locations or predictable trajectories.	Based on multi-antenna CSI available at the PHY layer. Can require additional AoA estimation algorithms.	Difficult to infer from eavesdropping. Could be inferred by raytracing and reproduced by choosing similar AoA.

2.3.2 Integration of PLA into Next-Generation Systems

This section discusses various aspects of integrating PLA schemes into next-generation wireless communication protocols. The discussion is mainly centered around channel-based PLA schemes in multiple-antenna systems, since this is the setup considered throughout this thesis. To clarify the discussion, we distinguish three types of deployment/channel properties:

- Slow/fast channel variations.
- Low/high transmitter mobility.
- Long/short range communication.

Protocol Aspects In a quickly changing wireless environment, channel-based PLA will require tracking of legitimate channels over time. This would only be feasible under protocols that guarantee frequent-enough transmissions such that accurate channel-tracking is possible. Such schemes would therefore be most relevant for protocols that are by design exchanging periodic transmissions, which in fact is a characterizing property of many mission-critical applications. On the other hand, for applications based on sporadic transmissions, PLA is not likely to be useful under fast-changing channel conditions. For such applications, channel-based PLA would rather be appropriate for short-range scenarios with low transmitter mobility, where stable features over time can be guaranteed.

As discussed in Section 2.2.1, feature-based PLA requires methods for feature tracking or learning which need to be taken into consideration in PLA-integrated protocol design. For features unknown prior to communication, initial trusted communication exchanges are required to obtain feature information. Initial trust must be established based on other methods (e.g., cryptographic authentication), which is an argument for why PLA can never entirely replace cryptographic schemes. For slowly time-variant features, re-calibration of legitimate features over time might be necessary, and for quickly time-varying features, feature-tracking algorithms are appropriate. To enable such schemes, one can envision a protocol that includes crypto-authentication every Nth packet, while $N-1$ packets are feature-tracked and authenticated using PLA.

PLA-integrated protocol design also need to handle rejected packets appropriately. A packet that triggers the PLA scheme could, from the perspective of the receiver, be either a false alarm or an actual attack. A straightforward protocol would simply drop the suspected packet at the PHY layer (i.e., not forward it to higher layers). Alternatively, decisions can be reported back to the transmitter with a request for a re-transmission. Whether decision-feedback and re-transmissions are necessary would in practice depend on how quickly information ages in the given application. In addition, more sophisticated protocols might integrate PHY layer decisions with higher layer information as part of a larger fault/intrusion detection system.

Deployment Aspects In short- to medium-range communications (around 10-50 m) with strong line-of-sight paths (e.g., remote control in an indoor factory location), the frequency-selectivity of the multi-path channel is typically low, which makes authentication based on frequency or impulse responses problematic. On the other hand, significant line-of-sight or reflective paths provide better conditions for predictable AoA features. Hence, location- and AoA-based beam-forming PLA appears as the most promising solution for short- to medium-range scenarios. These are the underlying conditions assumed throughout most of the papers in this thesis. In the opposite way, long-range communications are less likely to provide line-of-sight paths, but more likely to possess frequency-selectivity due to a richer multi-path environment. For this reason, frequency responses and multi-carrier CSI appear more promising for long-range communication scenarios.

The type of wireless deployment will highly influence the systems ability to effectively use channel-based PLA. In certain scenarios, it might even be appropriate to influence the deployment for security purposes (e.g., place sensors such that RSSI or AoA easily distinguishes the different sensors). Deployment design for security purposes has not been discussed in previous works on PLA. However, as many of the results provided in this thesis show (in particular under the optimal attack strategies in Paper D and F), choice of antenna deployments has a large impact on the detection performance of the PLA schemes.

Security and Reliability Aspects As outlined in Section 2.1, reliability forms the utmost priority in mission-critical and URLLC systems. The integration of PLA will have impacts on the system-level reliability due to the occasional false alarms and missed detections. False alarms causing packet drops or re-transmissions (depending on the protocol design) will inevitably lower the reliability and/or increase the latency. However, given that such performance impacts can be upper bounded, the impacts can be traded against the enhanced security from the PLA scheme. Such delay performance bounds is one of the contributions of this thesis.

The extent to which PLA could replace existing authentication methods is still a matter of future research and development. PLA could be viewed as an additional first line of defense that maintains security if session keys used for lightweight crypto-based authentication would leak to an adversary. Alternatively, PLA could maintain continuous authentication in a low-latency fashion until an attack is suspected and higher-layer authentication is invoked, or used together with periodic higher-layer authentication of a subset of transmissions.

Chapter 3

Preliminary Concepts

This chapter introduces the theoretical concepts and frameworks that are used throughout the rest of this thesis, here discussed with a greater level of detail than what was allowed in the appended publications. We begin by introducing binary hypothesis testing which is central to the PLA schemes studied in this thesis. Next, we cover some mathematical tools for analyzing distributions of complex Gaussian quadratic forms that are used for analyzing the detection performance of the PLA schemes. Finally, we introduce the stochastic network calculus framework that is used for deriving the delay performance bounds.

3.1 Hypothesis Testing

In this section, we provide the hypothesis testing framework commonly used for feature-based PLA under complex Gaussian feature vectors. The complex Gaussian feature distribution is relevant for many physical layer features arising in wireless communication systems, such as multi-carrier frequency responses, impulse responses, and multiple-antenna channels. Since the feature here generally represent a communication channel, commonly denoted with the symbol $\mathbf{h}$, we will use the notation $\mathbf{h}$[1] to denote a $(N \times 1)$ complex Gaussian feature vector in the following.

The feature-based PLA is based on the feature observation $\mathbf{h}$, following the model

$$\mathbf{h} \sim \begin{cases} \mathcal{CN}(\boldsymbol{\mu}_A, \boldsymbol{\Sigma}_A) & \text{if} \quad \mathcal{H}_0 \\ \mathcal{CN}(\boldsymbol{\mu}_E, \boldsymbol{\Sigma}_E) & \text{if} \quad \mathcal{H}_1, \end{cases} \tag{3.1}$$

where $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ respectively represent the mean and covariance matrix for Alice's $(i = A)$ and Eve's $(i = E)$ feature. Moreover, as already introduced in Chapter 2.2, $\mathcal{H}_0$ represents the legitimate hypothesis (i.e., that Alice is the source of the transmissions) and $\mathcal{H}_1$ represents the alternative hypothesis that the transmission is an

[1] In difference to the general feature vector $\mathbf{x}$ introduced in Chapter 2.2.

impersonation attempt by Eve. Further recapitulating from Chapter 2.2, except here without time indexing, the PLA procedure is designed as a hypothesis test

$$d(\mathbf{h}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} T, \tag{3.2}$$

where the problem is to choose the appropriate discriminant function $d(\mathbf{h})$.

3.1.1 Neyman-Pearson Test

For a standard Neyman-Pearson test (NPT), we need to assume that the channel distributions of the legitimate user and the attacker are completely known. Denote by $p_A(\mathbf{h})$ the likelihood of the observation given $\mathcal{H}_0$ and by $p_E(\mathbf{h})$ the likelihood of the observation given $\mathcal{H}_1$. According to the the Neyman-Pearson Lemma, the hypothesis test

$$\Lambda(\mathbf{h}) = \frac{p_E(\mathbf{h})}{p_A(\mathbf{h})} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \eta \tag{3.3}$$

is the most powerful test at a significance level (false alarm probability) $\alpha = \mathbb{P}(\Lambda(\mathbf{h}) > \eta|\mathcal{H}_0)$. This means that this is the test that minimizes the missed detection probability $\beta = \mathbb{P}(\Lambda(\mathbf{h}) < \eta|\mathcal{H}_1)$ for a given significance level α. Next, for the complex Gaussian densities, note that

$$\begin{aligned}\log\left(\frac{p_E(\mathbf{h})}{p_A(\mathbf{h})}\right) = \log\left(\frac{|\mathbf{\Sigma}_E|}{|\mathbf{\Sigma}_A|}\right) + \\ (\mathbf{h}-\boldsymbol{\mu}_A)^\dagger\mathbf{\Sigma}_A^{-1}(\mathbf{h}-\boldsymbol{\mu}_A) - (\mathbf{h}-\boldsymbol{\mu}_E)^\dagger\mathbf{\Sigma}_E^{-1}(\mathbf{h}-\boldsymbol{\mu}_E).\end{aligned} \tag{3.4}$$

Now by defining the squared Mahalanobis distances $D(\mathbf{h}\|p_A) = (\mathbf{h}-\boldsymbol{\mu}_A)^\dagger\mathbf{\Sigma}_A^{-1}(\mathbf{h}-\boldsymbol{\mu}_A)$ and $D(\mathbf{h}\|p_E) = (\mathbf{h}-\boldsymbol{\mu}_E)^\dagger\mathbf{\Sigma}_E^{-1}(\mathbf{h}-\boldsymbol{\mu}_E)$ and taking logarithm on both sides of (3.3), we get the NPT in the form

$$D(\mathbf{h}\|p_A) - D(\mathbf{h}\|p_E) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \eta', \tag{3.5}$$

where $\eta' = \log(\eta) - \log\left(\frac{|\mathbf{\Sigma}_E|}{|\mathbf{\Sigma}_A|}\right)$. This shows that the optimal Neyman-Person PLA test is to compare the difference in Mahalanobis distances to the threshold η'.

3.1.2 Composite Test (GLRT)

In the often more realistic scenario that the attacker's channel distribution is unknown to the receiver, and the Neyman-Person test clearly cannot be performed, then a composite hypothesis test is constructed instead. For simplicity, we can, without loss of information, define the test statistic $\mathbf{z} = \mathbf{h} - \boldsymbol{\mu}_A$ and re-define the hypotheses as follows

$$\begin{aligned}\mathcal{H}_0 &: \mathbf{z} \sim \mathcal{CN}(0, \boldsymbol{\Sigma}) \\ \mathcal{H}_1 &: \mathbf{z} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \text{ with } \boldsymbol{\mu} \neq 0.\end{aligned} \tag{3.6}$$

A hypothesis test which includes estimates of unknown parameters is called a generalized likelihood ratio test (GLRT). The GLRT for the setup in (3.6) is defined as

$$\Lambda(\mathbf{z}) = \frac{p(\mathbf{z}|\mathcal{H}_1, \hat{\boldsymbol{\mu}})}{p(\mathbf{z}|\mathcal{H}_0)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \eta, \tag{3.7}$$

where $\hat{\boldsymbol{\mu}}$ is some estimate of the mean $\boldsymbol{\mu}$ and η is a design threshold. Since no information on $\boldsymbol{\mu}$ is available, the best estimate of $\boldsymbol{\mu}$ conditioned on $\mathcal{H}_1$ is $\hat{\boldsymbol{\mu}} = \mathbf{z}$. Taking the log-likelihood yields

$$\begin{aligned}\log\left(\Lambda(\mathbf{z})\right) = \log\left(\left.\frac{\frac{1}{\pi^N|\boldsymbol{\Sigma}|}e^{-(\mathbf{z}-\hat{\boldsymbol{\mu}})^\dagger\boldsymbol{\Sigma}^{-1}(\mathbf{z}-\hat{\boldsymbol{\mu}})}}{\frac{1}{\pi^N|\boldsymbol{\Sigma}|}e^{-\mathbf{z}^\dagger\boldsymbol{\Sigma}^{-1}\mathbf{z}}}\right)\right|_{\hat{\boldsymbol{\mu}}=\mathbf{z}} = \\ \mathbf{z}^\dagger\boldsymbol{\Sigma}^{-1}\mathbf{z} = (\mathbf{h}-\boldsymbol{\mu}_A)^\dagger\boldsymbol{\Sigma}^{-1}(\mathbf{h}-\boldsymbol{\mu}_A) = D(\mathbf{h}\|p_A).\end{aligned} \tag{3.8}$$

Hence, the GLRT is given by

$$D(\mathbf{h}\|p_A) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \eta', \tag{3.9}$$

where $\eta' = \log(\eta)$ again is a design threshold. The geometrical interpretation is that the message is accepted if the Mahalanobis distance to the legitimate distribution $D(\mathbf{h}\|p_A)$ is below the threshold η'.

3.1.3 GLRT Error Probabilities

In this section, we provide the false-alarm and missed-detection probabilities under the GLRT hypothesis test (3.9). First, we establish the following remark:

Remark 3.1. *The GLRT test is often formulated as*

$$d(\mathbf{h}) = 2(\mathbf{h}-\boldsymbol{\mu}_A)^\dagger\boldsymbol{\Sigma}^{-1}(\mathbf{h}-\boldsymbol{\mu}_A) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} T, \tag{3.10}$$

with $T = 2\eta'$. Note that the outcomes of test (3.9) *and* (3.10) *are equivalent; however, the representation* (3.10) *allows for closed-form distribution of the discriminant function $d(\mathbf{h}) = 2D(\mathbf{h}\|p_A)$ in the lefthand side, which simplifies error analysis. For this reason, we will work with representation* (3.10) *throughout the rest of this section.*

First, we provide the false alarm probability, defined according to

$$p_{\text{FA}}(T) = \mathbb{P}(d(\mathbf{h}) > T|\mathcal{H}_0), \tag{3.11}$$

in the following lemma:

Lemma 3.1 (GLRT False Alarm Probability)**.** *The false alarm probability under the GLRT hypothesis test* (3.10) *is given by*

$$p_{FA}(T) = 1 - F_{\chi^2_{2N}}(T) \tag{3.12}$$

where $F_{\chi^2_{2N}}(T)$ denotes the cumulative distribution function of a central χ^2 random variable with $2N$ degrees of freedom.

Proof. Consider a Cholesky decomposition of the covariance matrix as $\mathbf{\Sigma} = \mathbf{C}^\dagger\mathbf{C}$. Now note that under $\mathcal{H}_0$ we have $\mathbf{h} \sim \mathcal{CN}(\boldsymbol{\mu}_A, \mathbf{\Sigma})$, and hence, we can write $d(\mathbf{h}) = \mathbf{y}^\dagger\mathbf{y}$ where

$$\mathbf{y} = \sqrt{2}(\mathbf{C}^\dagger)^{-1}(\mathbf{h} - \boldsymbol{\mu}_A) \underset{\mathcal{H}_0}{\sim} \mathcal{CN}(0, 2\mathbf{I}), \tag{3.13}$$

i.e., $\mathbf{y}$ is a $(N \times 1)$ vector with i.i.d circularly symmetric complex Gaussian entries. Consider forming the $(2N \times 1)$ vector $\mathbf{v}$ as

$$\mathbf{v} = \begin{bmatrix} \Re(\mathbf{y}) \\ \Im(\mathbf{y}) \end{bmatrix} \tag{3.14}$$

and note that $\mathbf{v}$ is a zero mean Gaussian with i.i.d standard variance entries. Hence, we have

$$d(\mathbf{h}) = \mathbf{y}^\dagger\mathbf{y} = \mathbf{v}^T\mathbf{v} = \sum_{n=1}^{2N} |v_n|^2, \tag{3.15}$$

from which follows, by definition, that $d(\mathbf{h}) \sim \chi^2_{2N}$. Finally, the false alarm probability (3.11) is the complementary CDF of the χ^2_{2N} distributed discriminant function, from which (3.12) follow.

□

Now we turn to the corresponding missed detection probability, defined according to

$$p_{\text{MD}}(T) = \mathbb{P}(d(\mathbf{h}) < T|\mathcal{H}_1), \tag{3.16}$$

and provided in the next lemma:

Lemma 3.2 (GLRT Missed Detection Probability)**.** *The missed detection probability under the GLRT hypothesis test* (3.10) *is given by*

$$p_{MD}(T) = F_{\chi^2_{2N}(\nu)}(T) \tag{3.17}$$

where $F_{\chi^2_{2N}(\nu)}(\cdot)$ denotes the cumulative distribution function of a non-central χ^2 random variable with $2N$ degrees of freedom and non-centrality parameter

$$\nu = 2(\boldsymbol{\mu}_E - \boldsymbol{\mu}_A)^\dagger\mathbf{\Sigma}^{-1}(\boldsymbol{\mu}_E - \boldsymbol{\mu}_A) \tag{3.18}$$

Proof. The proof is similar to that of Lemma 3.1. Again, consider the Cholesky decomposition $\mathbf{\Sigma} = \mathbf{C}^\dagger\mathbf{C}$. Under $\mathcal{H}_1$, we have $\mathbf{h} \sim \mathcal{CN}(\boldsymbol{\mu}_E, \mathbf{\Sigma})$. We rewrite $d(\mathbf{h}) = \mathbf{y}^\dagger\mathbf{y}$ with

$$\mathbf{y} = \sqrt{2}(\mathbf{C}^\dagger)^{-1}(\mathbf{h} - \boldsymbol{\mu}_A) \underset{\mathcal{H}_0}{\sim} \mathcal{CN}(\boldsymbol{\mu}_y, 2\mathbf{I}), \tag{3.19}$$

with $\boldsymbol{\mu}_y = \sqrt{2}(\mathbf{C}^\dagger)^{-1}(\boldsymbol{\mu}_E - \boldsymbol{\mu}_A)$. Following the same procedure as in (3.14), we can again create a $(2N \times 1)$ real-valued vector $\mathbf{u} \sim \mathcal{N}(\boldsymbol{\mu}_u, \mathbf{I})$ from the real and imaginary elements of $\mathbf{y}$. This allows us to write

$$d(\mathbf{h}) = \mathbf{y}^\dagger\mathbf{y} = \mathbf{u}^T\mathbf{u} = \sum_{n=1}^{2N} |u_n|^2, \tag{3.20}$$

and u_n are independent standard variance Gaussian random variables with mean μ_n. This implies, by definition, that $d(\mathbf{h}) \sim \chi^2_{2N}(\nu)$, with non-centrality parameter

$$\nu = \boldsymbol{\mu}_u^T\boldsymbol{\mu}_u = \boldsymbol{\mu}_y^\dagger\boldsymbol{\mu}_y = 2(\boldsymbol{\mu}_E - \boldsymbol{\mu}_A)^\dagger\mathbf{\Sigma}^{-1}(\boldsymbol{\mu}_E - \boldsymbol{\mu}_A). \tag{3.21}$$

Finally, the probability of missed detection in (3.29) is the CDF of the non-central χ^2 distributed discriminant function, from which expression (3.17) follows.

□

Remark 3.2. *We note that the missed detection probability in Lemma 3.2 is predicated on the assumption that $\mathbf{\Sigma}_E = \mathbf{\Sigma}_A = \mathbf{\Sigma}$. This can easily be generalized to the case of $\mathbf{\Sigma}_E = \alpha\mathbf{\Sigma}_A$ for some constant α, as we for instance will show in Paper B.*

3.2 Distributions of Complex Gaussian Quadratic Forms

As shown in the previous section, GLRT hypothesis testing based on complex Gaussian features often results in decisions based on quadratic forms in complex Gaussian vectors. Note that the missed detection probability in Lemma 3.2 constitutes a special case where a closed-form distribution is possible. However, in this thesis we encounter several variations of this type of problem where closed form expressions are not tractable or more complicated to derive. For the purpose of demonstrating the different tools used for analyzing distributions of complex quadratic forms, let us in this section consider the generic complex quadratic form

$$Y = \mathbf{x}^\dagger\mathbf{A}\mathbf{x}, \tag{3.22}$$

where $\mathbf{x} \sim \mathcal{CN}(\boldsymbol{\mu}, \mathbf{\Sigma})$ is a $(1 \times N)$ CSCG vector with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma}$, and $\mathbf{A}$ is a $(N \times N)$ matrix. The problem we are interested in is finding an expression for the cumulative distribution

$$F_Y(y) = \mathbb{P}(Y < y). \tag{3.23}$$

Generally, (3.23) is not tractable in closed form (apart from in certain special cases as shown below) but there exist efficient approximation techniques in the previous literature.

In this thesis, we apply several different approximation techniques for evaluating (3.23). We will start by presenting tools that apply under the assumption of a symmetric and positive semidefinite matrix $\mathbf{A}$.

3.2.1 Positive Semidefinite Quadratic Forms

Note that through an appropriate choice of linear transformations the quadratic form (3.22) can be expanded into a weighted sum of independent non-central chi squared random variables. This is summarized in the following lemma:

Lemma 3.3 (Weighted Sum of Non-Central Chi-Squared Random Variables)**.** *The quadratic form can be decomposed into*

$$Y = \sum_{i=1}^{N} \lambda_i Y_i, \tag{3.24}$$

where λ_i are the eigenvalues of $\mathbf{\Sigma A}$ and Y_i $\chi_2^2(\nu_i)$ are independent non-central chi-squared random variables with two degrees of freedom and non-centrality parameters ν_i.

Proof. Consider the linear transformation $\mathbf{x} = \mathbf{Q}\bar{\mathbf{x}}$ where $\mathbf{Q}$ is defined by the Cholesky decomposition of the covariance matrix $\mathbf{\Sigma} = \mathbf{Q}^\dagger\mathbf{Q}$. This transformation allows us to write

$$Y = \bar{\mathbf{x}}^\dagger \mathbf{Q}^\dagger \mathbf{A} \mathbf{Q} \bar{\mathbf{x}}, \tag{3.25}$$

where we have $\bar{\mathbf{x}} \sim \mathcal{CN}(\mathbf{Q}^{-1}\boldsymbol{\mu}, \mathbf{I})$. Considering the eigenvalue decomposition $\mathbf{Q}^\dagger\mathbf{A}\mathbf{Q} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\dagger$ and defining $\mathbf{y} = \mathbf{U}^\dagger\bar{\mathbf{x}}$ results in

$$Y = \mathbf{y}^\dagger \mathbf{\Lambda} \mathbf{y} = \sum_{i=1}^{N} \lambda_i |y_i|^2 \tag{3.26}$$

where $\lambda_i > 0$ are the non-negative eigenvalues of $\mathbf{\Sigma A}$ (and diagonal elements of $\mathbf{\Lambda}$), and y_i denote the elements of the transformed vector $\mathbf{y} = \mathbf{U}^\dagger\mathbf{Q}^{-1}\mathbf{x}$. □

Firstly, note that if all eigenvalues are equal we can obtain the CDF in closed form, as shown in the following lemma:

Lemma 3.4 (Distribution for Equal Eigenvalues)**.** *Under the assumption of equal weights $\lambda_1 =, \cdots, \lambda_N = \lambda$, the distribution can be given in closed form according to*

$$F_Y(y) = F_{\chi^2_{2N}(\nu)}(y/\lambda), \tag{3.27}$$

i.e., a non-central chi-squared distribution with non-centrality parameter $\nu = \sum_{i=1}^{N} \nu_i$.

Proof. Under the assumption $\lambda_i = \lambda$ for $i = 1, \cdots, N$, we get from (3.26) that

$$Y = \lambda \sum_{i=1}^{N} |y_i|^2 = \lambda \bar{Y}. \tag{3.28}$$

The sum $\bar{Y}$ is the sum of independent $\chi^2_2(\nu_i)$ random variables, from which follows that $\bar{Y} \sim \chi^2_{2N}(\nu)$. □

For the general case with arbitrary positive eigenvalues, [86] have provided a series expansion of the CDF (3.23). This series expansion, which is used for the missed detection probability for a distributed PLA receiver in Paper C of this thesis, is provided in the following theorem:

Theorem 3.1 (Series Expansion for Positive Semidefinite Quadratic Forms)**.** *The CDF in* (3.23) *can be obtained as*

$$F_Y(y) = \sum_{i=0}^{\infty} c_i F_{2(i+\bar{k})}(y/\beta) \tag{3.29}$$

for any $0 < \beta \leq \min(\lambda_1, \cdots, \lambda_N)$ *with* $\bar{k} = \sum_{i=0}^{N} k_i$, $\kappa_i = 1 - \beta/\lambda_i$,

$$g_k = \sum_{i=0}^{N} \kappa_i^k + k/2 \sum_{i=0}^{N} \nu_i \kappa_i^{k-1}(1 - \kappa_i), \tag{3.30}$$

$$c_0 = \prod_{i=0}^{N} (\beta/\lambda_i) \exp\left(-1/2 \sum_{i=0}^{N} \nu_i\right), \tag{3.31}$$

and $c_k = k^{-1} \sum_{r=0}^{k-1} g_{k-r} c_r$ *for* $k \geq 1$.

Proof. From Lemma 3.3, we know that $F_Y(y) = \mathbb{P}(\sum_{j=1}^{N} \lambda_j Y_j < y)$ with Y_j distributed according to $\chi^2_{k_i}(\nu_i)$. The series expansion (3.29) of CDFs of weighted sums of non-central chi-squared random variables is provided in [86, Section VI]. □

Remark 3.3. *Note that Theorem 3.1 presupposes distinct eigenvalues* $\lambda_1, \cdots, \lambda_N$. *In the case of eigenvalues with algebraic multiplicity* $m > 1$, *the terms in* (3.24) *corresponding to the same eigenvalue can be grouped to a single* χ^2 *with higher degree of freedom. The slightly more general version of the expansion provided in [86] allows for weighted sums with varying degrees of freedom.*

3.2.2 Indefinite Quadratic Forms (Saddle-Point Approximation)

The previously mentioned methods do not generalize well to the case of indefinite $\mathbf{A}$. When such problems are encountered, we will use a saddle-point approximation technique. The saddle-point approximation, which is originally provided by [87], can be summarized into the following steps:

1. Represent the CDF (3.23) as a single-dimensional complex integral.

2. Use a second order Taylor approximation of the integrand around a stationary point. This approximation allows a closed form expression for the integral in terms of the second order derivative.

To perform the first step, we will reuse the representation from (3.25) in Section 3.2.1

$$Y = \bar{\mathbf{x}}^\dagger \bar{A} \bar{\mathbf{x}}, \tag{3.32}$$

with $\bar{\mathbf{x}} \sim \mathcal{CN}(\mathbf{b}, \mathbf{I})$, $\mathbf{b} = \mathbf{Q}\boldsymbol{\mu}$, and $\bar{\mathbf{A}} = \mathbf{Q}^\dagger \mathbf{A} \mathbf{Q}$, i.e., the quadratic from represented in terms of the whitened vector $\bar{\mathbf{x}}$. Furthermore, we will use the eigenvalue decomposition $\bar{\mathbf{A}} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\dagger$. The first step is achieved by the following proposition:

Proposition 3.1 (Integral Representation of CDF)**.** *The CDF* (3.23) *can be expressed as*

$$F_Y(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{y(\beta+j\omega)-c(\omega)}}{(\beta+j\omega)|\mathbf{I}+(\beta+j\omega)\boldsymbol{\Lambda}|} d\omega, \tag{3.33}$$

with the arbitrary real constant $\beta > 0$, $\mathbf{b} = \mathbf{Q}_E^\dagger \boldsymbol{\mu}_E$, *and*

$$c(\omega) = \mathbf{b}^\dagger \left(\mathbf{I} + \frac{1}{\beta+j\omega} \boldsymbol{\Lambda}^{-1} \right)^{-1} \mathbf{b}. \tag{3.34}$$

Proof. Introducing, $\bar{\mathbf{x}} = \mathbf{b} + \mathbf{h}$, we can integrate over the PDF of $\mathbf{h} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ according to

$$F_Y(y) = \int_{-\infty}^{\infty} \frac{1}{\pi^N} e^{-\mathbf{h}^\dagger \mathbf{h}} \mathcal{U}\left(y - (\mathbf{b}+\mathbf{h})^\dagger \bar{\mathbf{A}} (\mathbf{b}+\mathbf{h})\right) d\mathbf{h}, \tag{3.35}$$

where $\mathcal{U}(\cdot)$ denotes the Heaviside step function. By using the Laplace representation $\mathcal{U}(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{x(\beta+j\omega)}}{\beta+j\omega} d\omega$ valid for $\beta > 0$, we can rewrite (3.35) as

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{1}{\pi^N} e^{-\mathbf{h}^\dagger \mathbf{h}} \left[\frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{(\beta+j\omega)\left(y-(\mathbf{b}+\mathbf{h})^\dagger \bar{\mathbf{A}}(\mathbf{b}+\mathbf{h})\right)}}{\beta+j\omega} d\omega \right] d\mathbf{h} = \\ \frac{1}{2\pi^{N+1}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{e^{y(\beta+j\omega)-\mathbf{h}^\dagger\mathbf{h}-(\mathbf{b}+\mathbf{h})^\dagger(\beta+j\omega)\bar{\mathbf{A}}(\mathbf{b}+\mathbf{h})}}{\beta+j\omega} d\omega d\mathbf{h}. \end{aligned} \tag{3.36}$$

Using the decomposition $\bar{\mathbf{A}} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\dagger$, the transformations $\tilde{\mathbf{h}} = \mathbf{U}^\dagger \mathbf{h}$ and $\bar{\mathbf{b}} = \mathbf{U}^\dagger \mathbf{b}$, and the fact that $d\mathbf{h} = d\tilde{\mathbf{h}}$, we can write (3.36) as

$$\frac{1}{2\pi^{N+1}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{e^{y(\beta+j\omega)-\tilde{\mathbf{h}}^\dagger\tilde{\mathbf{h}}-(\tilde{\mathbf{h}}+\bar{\mathbf{b}})^\dagger(\beta+j\omega)\mathbf{D}(\tilde{\mathbf{h}}+\bar{\mathbf{b}})}}{\beta+j\omega}d\omega d\tilde{\mathbf{h}} = \frac{1}{2\pi^{N+1}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{e^{y(\beta+j\omega)-(\tilde{\mathbf{h}}+\tilde{\mathbf{b}})^\dagger\mathbf{B}(\tilde{\mathbf{h}}+\tilde{\mathbf{b}})-c(\omega)}}{\beta+j\omega}d\omega d\tilde{\mathbf{h}}, \tag{3.37}$$

with $\mathbf{B}=\mathbf{I}+(\beta+j\omega)\mathbf{\Lambda}$, $\tilde{\mathbf{b}}=\left(\mathbf{I}+\frac{1}{\beta+j\omega}\mathbf{\Lambda}^{-1}\right)^{-1}\bar{\mathbf{b}}$, and

$$c(\omega)=\bar{\mathbf{b}}^\dagger\left(\mathbf{I}+\frac{1}{\beta+j\omega}\mathbf{\Lambda}^{-1}\right)^{-1}\bar{\mathbf{b}}. \tag{3.38}$$

We can then integrate out $\tilde{\mathbf{h}}$ by noting that (3.37) can be written as

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{\pi^N}e^{-(\tilde{\mathbf{h}}+\tilde{\mathbf{b}})^\dagger\mathbf{B}(\tilde{\mathbf{h}}+\tilde{\mathbf{b}})}d\tilde{\mathbf{h}}\frac{e^{y(\beta+j\omega)-c(\omega)}}{\beta+j\omega}d\omega = \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{y(\beta+j\omega)-c(\omega)}}{(\beta+j\omega)|\mathbf{I}+(\beta+j\omega)\mathbf{\Lambda}|}d\omega, \tag{3.39}$$

where we have used the fact that the Gaussian integral $\int_{-\infty}^{\infty}\frac{1}{\pi^N}e^{-(\tilde{\mathbf{h}}+\tilde{\mathbf{b}})^\dagger\mathbf{B}(\tilde{\mathbf{h}}+\tilde{\mathbf{b}})}d\tilde{\mathbf{h}}=\frac{1}{|\mathbf{B}|}$ is solvable in closed form. □

Although neither the simplified integral in the previous Proposition 3.1 is computable in closed form, it is easier to handle than the brute force N-dimensional integral over the complex vector $\mathbf{h}$. As summarized above, one can use a saddle-point approximation of the integral (3.33). This technique is presented in the following theorem:

Theorem 3.2 (Saddle-Point Approximation)**.** *The CDF* (3.23) *can be approximately evaluated as*

$$F_Y(y)\approx\frac{1}{2\pi}e^{s(z_0)}e^{-j\angle s''(z_0)}\sqrt{\frac{2\pi}{|s''(z_0)|}}, \tag{3.40}$$

where

$$s(z)=yz-\mathbf{b}^\dagger\left(\mathbf{I}+\frac{1}{z}\mathbf{\Lambda}^{-1}\right)^{-1}\mathbf{b}-\ln(z)-\ln(|\mathbf{I}+z\mathbf{\Lambda}|), \tag{3.41}$$

and z_0 is a stationary point such that $s'(z_0)=0$.

Proof. With a change to the complex variable $z=\beta+j\omega$, we can write (3.33) as

$$F_Y(y)=-\frac{1}{j2\pi}\oint_{-\beta-j\infty}^{-\beta+j\infty}e^{-s(z)}dz \tag{3.42}$$

with $s(z)$ defined according to (3.41). The saddle point method uses the approximation $s(z) \approx s(z_0) + \frac{1}{2}s''(z_0)(z-z_0)^2$ to write

$$\begin{aligned} F_Y(y) &\approx \frac{1}{j2\pi} \oint_{\beta-j\infty}^{\beta+j\infty} e^{-(s(z_0)+\frac{1}{2}s''(z_0)(z-z_0)^2)} dz \\ &= \frac{1}{j2\pi} e^{-s(z_0)} \oint_{\beta-j\infty}^{\beta+j\infty} e^{-\frac{1}{2}s''(z_0)(z-z_0)^2} dz = \frac{1}{j2\pi} e^{-s(z_0)} e^{j\phi} \sqrt{\frac{2\pi}{|s''(z_0)|}} \end{aligned} \tag{3.43}$$

with $\phi = \frac{\pi - \angle s''(z_0)}{2}$. Finally, we note that $e^{j\phi} = je^{-j\angle s''(z_0)}$ from which (3.40) follows. □

Some final remarks regarding the saddle-point approximation are in order:

1. The first and second order derivatives of $s(z)$ are needed for numerical evaluation of the approximation in Theorem 3.2. These can fairly straightforwardly be derived from (3.41) and, although left out here, they can be found in [87, Section V].

2. To the best of our knowledge, no closed form solutions for the stationary point $s'(z_0) = 0$ exist. The numerical results based on the saddle-point approximation are in this thesis obtained from numerical root-finding algorithms, while carefully compared against Monte Carlo simulations.

3. Encountering the complementary CDF (CCDF) $\mathbb{P}(Y > y)$, as for instance in Paper D of this thesis, it is more convenient to derive Proposition 3.1 based on the Laplace representation of $1 - \mathcal{U}(x) = -\frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{x(j\omega-\beta)}}{j\omega-\beta} d\omega$. The derivations of Proposition 3.1 and Theorem 3.2 are in this case very similar; however, we provide the corresponding proofs for the CCDF in Paper D.

3.3 Stochastic Network Calculus

Stochastic network calculus is a mathematical framework that allows us to analyze input-output relationships of stochastic queueing systems through, for example, performance bounds on delay or backlog given arrival and service distributions. In this thesis, we use tools from stochastic network calculus to bound the queueing delays introduced by PLA schemes. For a complete overview of stochastic network calculus, we refer to [88].

Queueing Model The queueing models in this thesis are based on the bivariate stochastic processes

$$A(\tau, t) = \sum_{k=\tau}^{t} a_k, \quad D(\tau, t) = \sum_{k=\tau}^{t} d_k,$$

representing the cumulative arrivals to and departures from a queue in the time interval $[\tau, t)$ for all $0 \leq \tau \leq t$. In time-slot k, a_k represents the instantaneous arrivals measured in bits, and d_k represent the instantaneous departures from the queue. The ability to transfer data from the buffer queue to the destination is characterized by the cumulative service process $S(\tau, t) = \sum_{k=\tau}^{t} s_k$. In this thesis, this queueing system generally represents a wireless communication link, where arrivals represent information to be transmitted (e.g., sensor measurements) and departures represent information successfully decoded at the receiver.

A widely used measure on the queueing system's ability to meet delay requirements is the *delay violation probability* [89]. The queueing delay at time point t is defined as

$$W(t) \triangleq \inf\{u > 0; A(0, t) \leq D(0, t + u)\}, \tag{3.44}$$

representing the frames required to serve the bits in the queue at time t. This delay is randomly varying due to the random service process and the delay violation probability is defined as $p(w) = \mathbb{P}(W(t) > w)$, i.e., the probability that a bit is not received within a defined deadline w. In many cases, an exact expression for the delay violation probability is complicated to derive. However, stochastic network calculus provides statistical bounds on this function.

Stochastic Network Calculus in the SNR Domain The work in [89] developed the stochastic network calculus framework for wireless fading links by observing that the analysis is simplified by converting the bivariate stochastic processes $A(\tau, t)$, $S(\tau, t)$ and $D(\tau, t)$ into $\mathcal{A}(\tau, t) \triangleq e^{A(\tau,t)}$, $\mathcal{S}(\tau, t) \triangleq e^{S(\tau,t)}$ and $\mathcal{D}(\tau, t) \triangleq e^{D(\tau,t)}$. This transformation allows the characterization of the random service process in terms of the varying instantaneous signal-to-noise ratio (SNR) due to fading of a wireless link. This is referred to as transforming the bit-domain processes into the SNR-domain since the service process, which is often logarithmic in the SNR in wireless systems, instead become linear in the instantaneous SNR. Arrival processes in the SNR-domain can then be seen as instantaneous SNR demands. In bit-domain, stochastic network calculus is based on a $(\min, +)$ dioid algebra over $\mathbb{R}^+$. Stochastic network calculus in the SNR-domain, on the other hand, is instead based on the $(\min, \times)$ dioid algebra since processes in the SNR-domain become multiplicative instead of additive. The performance bounds, which can be seen as variations of moment bounds, are derived in terms of Mellin transforms of the involved queueing processes. The Mellin transform of a random variable X, closely related to the moment-generating function (MGF), is defined as $\mathcal{M}_X(s) = \mathbb{E}[X^{s-1}]$.

Statistical Delay Bound The upper bound on the delay violation probability we utilize in this thesis is given by the following lemma:

Lemma 3.5. *For $s > 0$,*

$$p(w) \leq \mathcal{K}(s, t+w, t), \tag{3.45}$$

where $\mathcal{K}(s,\tau,t)$ *is called the kernel and given by*

$$\mathcal{K}(s,\tau,t) \triangleq \sum_{u=0}^{\min(\tau,t)} \mathcal{M}_{\mathcal{A}}(1+s,u,t)\mathcal{M}_{\mathcal{S}}(1-s,u,\tau), \tag{3.46}$$

and $\mathcal{M}_{\mathcal{S}}(s,\tau,t) = \mathbb{E}[\mathcal{S}(\tau,t)^{s-1}]$ *and* $\mathcal{M}_{\mathcal{A}}(s,\tau,t) = \mathbb{E}[\mathcal{A}(\tau,t)^{s-1}]$ *are Mellin transforms of the independent SNR-domain service and arrival processes.*

Proof. See Theorem 1 in [89]. □

The delay bound provided by Lemma 3.5 is particularly suitable for performance evaluation in mission-critical networks since it provides an upper limit on the delay violation probability, i.e., a real system operating under the assumed conditions will certainly achieve a better delay performance.

In the following, assuming i.i.d. instantaneous arrivals and service, we can write $\mathcal{M}_{\mathcal{S}}(s,\tau,t) = \mathcal{M}_{\mathcal{S}}(s)^{t-\tau}$ and $\mathcal{M}_{\mathcal{A}}(s,\tau,t) = \mathcal{M}_{\mathcal{A}}(s)^{t-\tau}$, where $\mathcal{M}_{\mathcal{S}}(s) \triangleq \mathbb{E}[e^{s_k(s-1)}]$ and $\mathcal{M}_{\mathcal{A}}(s) \triangleq \mathbb{E}[e^{a_k(s-1)}]$ due to the independence of the instantaneous service and arrivals s_k and a_k. Then, assuming stationarity of the underlying queueing processes, we let $t \to \infty$ in the righthand side of (3.45) and get

$$\lim_{t\to\infty} \mathcal{K}(s, t+w, t) = \frac{\mathcal{M}_{\mathcal{S}}(1-s)^w}{1 - \mathcal{M}_{\mathcal{A}}(1+s)\mathcal{M}_{\mathcal{S}}(1-s)}, \tag{3.47}$$

under the stability condition $\mathcal{M}_{\mathcal{A}}(1+s)\mathcal{M}_{\mathcal{S}}(1-s) < 1$ required for the sum in (3.46) to converge. Since Lemma 3.5 holds for all $s > 0$, it follows that minimization of (3.47) over $s > 0$ gives us an asymptotic upper bound on the delay violation probability. Hence, for the stable and stationary queueing system, the upper bound on the delay violation probability can be compactly written as

$$p(w) \leq \inf_{s>0} \left\{ \lim_{t\to\infty} \mathcal{K}(s, t+w, t) \right\} \tag{3.48}$$

with the objective function to be minimized given by the steady-state kernel in (3.47). This function can be shown to be a convex function for every s in the stability interval $\mathcal{M}_{\mathcal{A}}(1+s)\mathcal{M}_{\mathcal{S}}(1-s) < 1$ (see Theorem 1 in [90]).

Some final remarks regarding stochastic network calculus:

1. Generally, no analytical tools from convex optimization can be applied to the minimization of (3.47) over s. Therefore, numerical evaluations conducted in this thesis are based on numerical grid search for the minimization over s.

2. Derivation of the service-process Mellin transform is generally the central problem in the delay-analysis parts of this thesis. For the considered system models and architectures, we (i) develop appropriate models for the service

process $\mathcal{S}(\tau, t)$, (ii) derive the following Mellin transform $\mathcal{M}_{\mathcal{S}}(s)$, and (iii) utilize Lemma 3.5 for an upper bound on the delay violation probability.

3. Finally, it is worth noting that alternative stochastic network calculus approaches exist that may be used for this analysis including *effective capacity* [91] and *MGF-based analysis* [92]. Nevertheless, the usefulness of the approach in [90] that we employ is most apparent when applied to wireless fading channels as the Mellin transform $\mathcal{M}_{\mathcal{S}}$ is already derived for many fading channels in the literature, e.g., [93–95]. This makes the approach particularly attractive for wireless networks analysis.

Part II

Included Papers